Mathematical Models for protein polymerization and applications to amyloid diseases

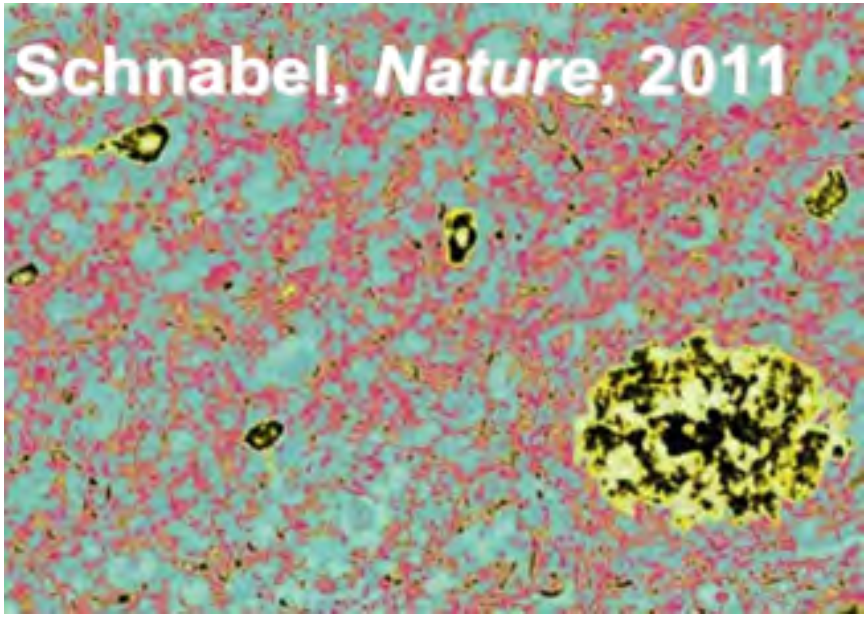
Marie Doumic

Taipei, May 27th, 2014

Protein polymerization

Common point between:

- Alzheimer's (illustrated)
- Prion (mad cow)
- Huntington's
- and some others (Parkinson's, etc)?



Neurodegenerative diseases characterized by abnormal accumulation of protein aggregates called AMYLOIDS

Prions: PRoteinaceous Infectious ONly particle

Healthy state: **mono**meric protein (PrP Prion, Aβ Alzheimer's, PolyQ Huntington's)

Disease state: polymers. PRION: PRoteinasceous Infectious Only
Prion: infectious agents composed of proteins in a misfolded form.

1

[1] Source:"Mad Cow Mysteries", a public lecture series given by Jay Ingram.

Protein polymerization: main questions

- Understand what are the key polymerization mechanisms
- Identify transient species
- What are the "most infectious" sizes of polymers?
- How to select and calibrate the models?

Main difficulty: **extremely rare and partial** size distributions

Models Derivation

- discrete (ODE) models

Models Derivation

- discrete (ODE) models
- continuous (PDE) models

Models Derivation

- discrete (ODE) models
- continuous (PDE) models
- ... and later on: stochastic models

Models for protein polymerization

Deterministic framework: the law of mass action (Guldberg & Waage, 1867)

$$A + B \underset{k^-}{\overset{k^+}{\rightleftharpoons}} A' + B'$$

is translated by

$$\frac{d[A]}{dt} = -k^+[A][B] + k^-[A'][B'],$$

and similarly for the concentrations $[B]$, $[A']$ and $[B']$.
Large amount of reactants

Models for protein polymerization

Reactions

V^* monomers' concentration, n_i $i-$polymer's concentration.

- monomer-conformer exchange $V^* \underset{k_l^-}{\overset{k_l^+}{\rightleftharpoons}} V$
- nucleation $\underbrace{V + \cdots + V}_{i_0} \underset{k_{off}^N}{\overset{k_{on}^N}{\rightleftharpoons}} n_{i_0}$,
- (de)polymerization by monomer addition $n_i + V \underset{g_{i+1}^-}{\overset{g_i}{\rightleftharpoons}} n_{i+1}$
- fragmentation-coalescence: $n_i + n_j \underset{B_{i+j} k_{i+j,i}}{\overset{k_{i,j}^{col}}{\rightleftharpoons}} n_{i+j}$
- etc.

Models for protein polymerization

ODE system for the "Prion" Model (Masel *et al.*, Biophys. Chem., 1999)

Considered reactions: polymerization and fragmentation

$$\frac{dn_i}{dt} = -V(t)\left(g_i n_i - g_{i-1} n_{i-1}\right) - B_i n_i + 2 \sum_{j=i+1}^{\infty} B_j k_{i,j} n_j, \qquad i \geq i_0,$$

$$\frac{dV}{dt} = \lambda - \gamma V - V \sum_{i=i_0}^{\infty} g_i n_i + 2 \sum_{j \geq i_0} \sum_{i < i_0} i k_{i,j} B_j n_j.$$

Mass balance:

$$\frac{d}{dt}\left(V(t) + \sum_{i=i_0}^{\infty} i n_i(t) \right) = \lambda - \gamma V.$$

Amyloid diseases: average size $i_M \gg 1$

Protein polymerization: from ODE to PDE

Defining

$$\varepsilon = \frac{1}{i_M}, \qquad x_i = i\varepsilon$$

with i_M the average polymer size:

$$i_M = \frac{\sum_{i \geq i_0} i n_i}{\sum n_i}$$

we may define, after definition of dimensionless quantities

$$n^\varepsilon(t,x) = \sum n_i \mathbb{1}_{[x_i, x_{i+1}]}$$

and get a PDE to replace the infinite ODE system.
Limit of such a derivation: the average size changes with time.

Protein polymerization: from ODE to PDE

- Coagulation & fragmentation:
  P. Laurençot & S. Mischler, Proc. Roy. Soc. Edinburgh, 2002
- polymerization & depolymerization in a closed system:
  From Bekker-Döring to Lifshitz-Slyozov: J.F. Collet, T. Goudon, F. Poupaud, A. Vasseur, SIAM App. Math, 2002
  2nd order approx: modified Lifshitz-Slyozov: Collet and Hariz, 2002.
- "Prion Model": MD, T. Goudon, T. Lepoutre, CMS 2009
- General Model with nucleation: S. Prigent et al, Plos One 2012.

Protein polymerization: from ODE to PDE
Formal derivation of a general model

Model proposed in (S. Prigent et al., Plos One, 2012):

$$\frac{dV^*}{dt} = -k_I^+ V^* + k_I^- V,$$

$$\frac{dV}{dt} = k_I^+ V^* - k_I^- V - \frac{i_0\, k_{on}^N\, V^{i_0+1} g(x_0)}{k_{off}^N + g(x_0)V} - V \int_{x_0}^{\infty} g\, n dx + \int_{x_0}^{\infty} g^- n dx,$$

$$\frac{\partial n}{\partial t} = -V \frac{\partial}{\partial x}(gn) + \frac{\partial}{\partial x}(g^- n) + 2\int_x^{\infty} B(y) k(x,y)\, n(t,y) dy - Bn$$
$$+\frac{1}{2}\int_{x_0}^{x} k_{col}(y, x-y) n(t,y) n(t, x-y) dy - \int_{x_0}^{\infty} k_{col}(x,y) n(t,x)\, n(t,y) dy,$$

$$g(x_0) n(t, x_0) = g(x_0) \frac{k_{on}^N\, V^{i_0}}{k_{off}^N + g(x_0)V}.$$

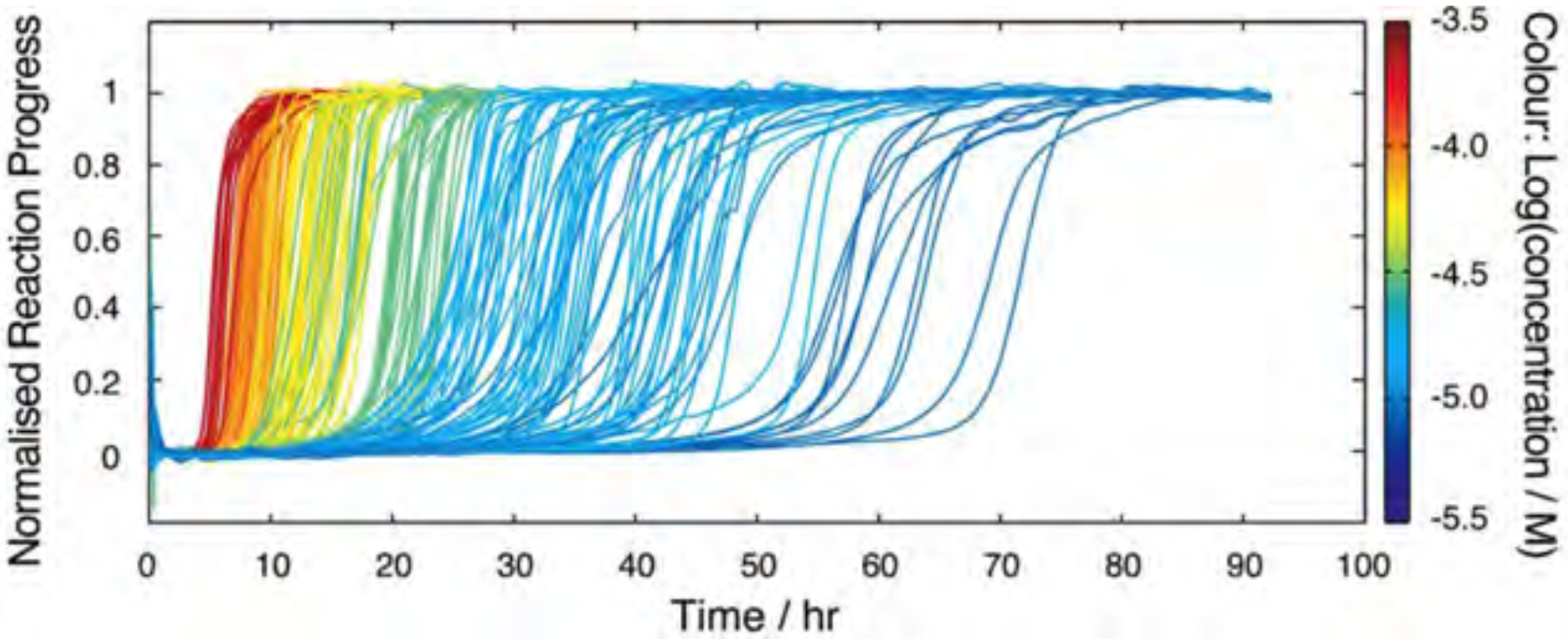
Towards models calibration: what can we measure?

Protein polymerization: measurements

1. ThT "Thioflavine T"

Measures the total polymerised mass

$$\sum_{i \geq i_0} i n_i, \qquad \int x n(t,x) dx$$



In vitro polymerization of $\beta 2m$. From Xue, Radford et al., PNAS, 2008

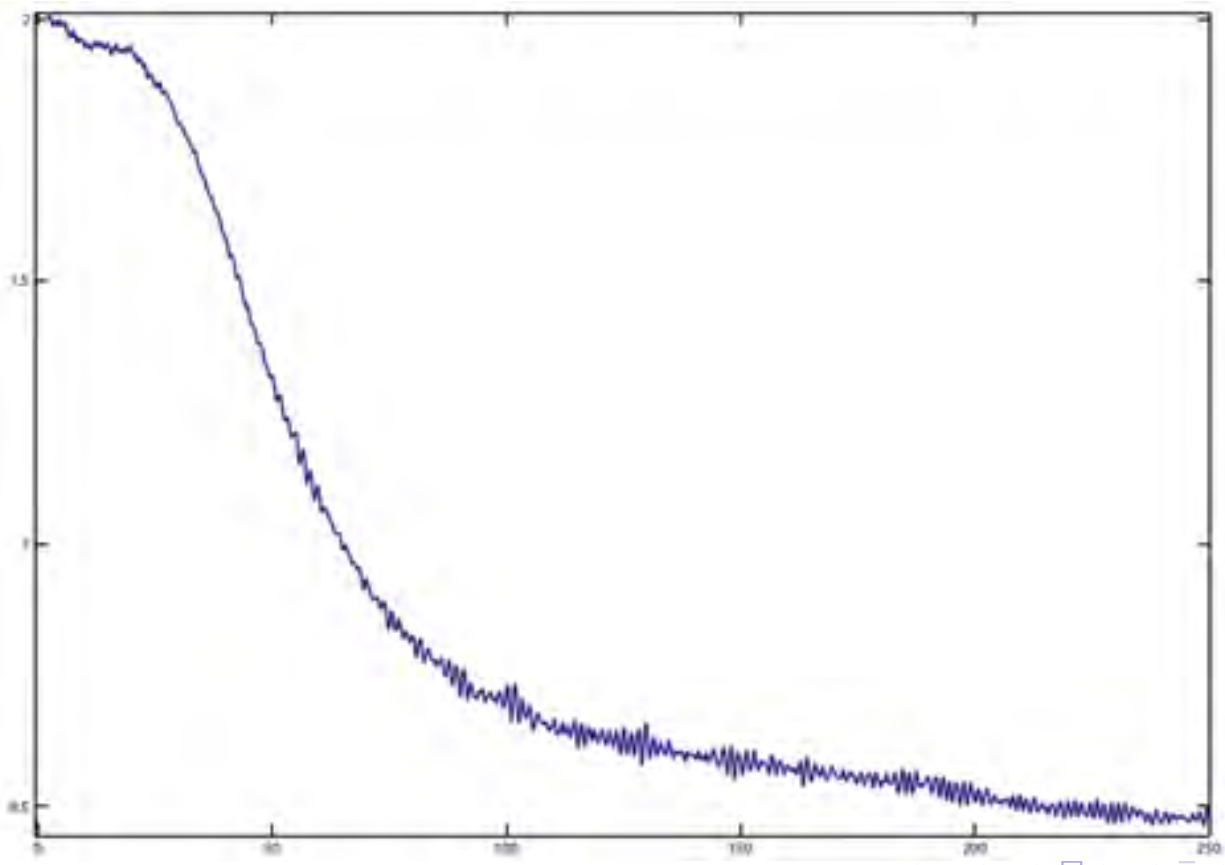
Protein polymerization: measurements

2. SLS, "Static Light Scattering"

Measures the second momentum

$$\sum_{i \geq i_0} i^2 n_i, \qquad \int x^2 n(t,x) dx$$

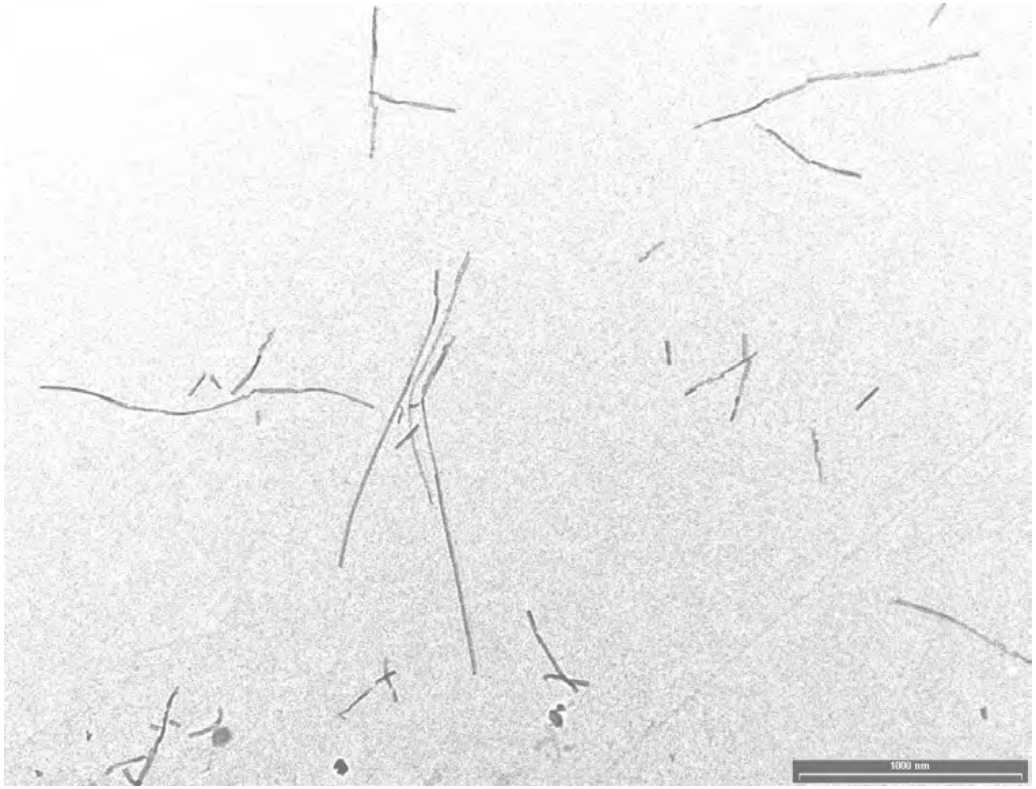
Depolymerization experiment (PrP)



Protein polymerization: measurements
3/ Size distributions

Difficult to obtain: image analysis + size threshold (measures only for more than 120 mers) + experimental device

PrP Misfolding (S. Prigent)



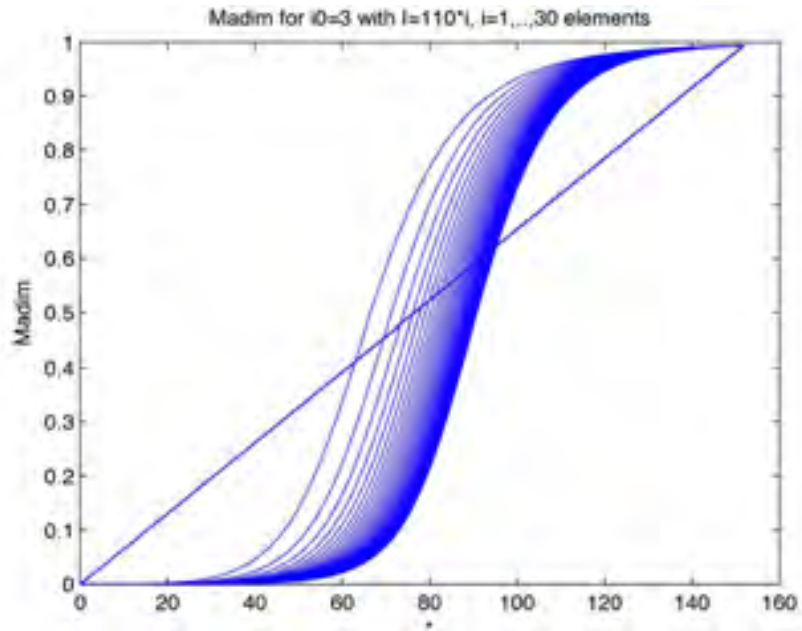
Towards models calibration: build efficient numerical schemes

Numerical Schemes

Very large computation domain (range 10^6)
Efficiency of the PDE for intensive numerical simulations

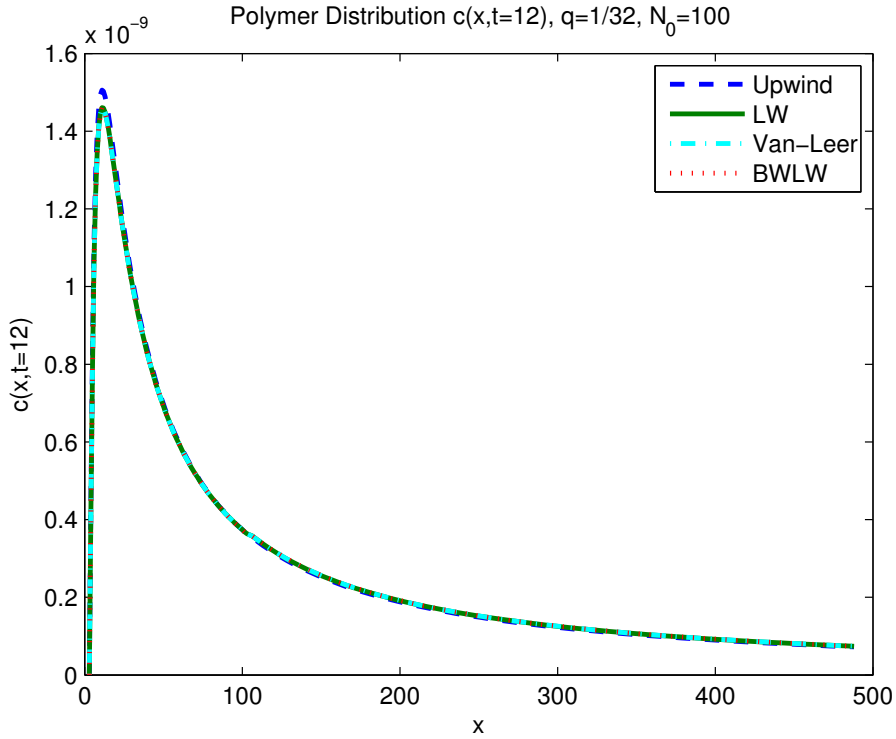
BUT

- No quantitative convergence rate for the (weak *) limit
- Specific treatment for the boundary condition / small sizes



Numerical Schemes

(H.T. Banks, MD, C. Kruse, submitted)



- ODE solved for $i \leq N_0 = O(\frac{1}{\varepsilon})$
- PDE is approximated e.g. by finite volume methods (upwind, Lax-Wendroff, flux limiter)

Numerical Schemes

(H.T. Banks, MD, C. Kruse, submitted)

- Adaptive mesh, refined towards smaller polymer sizes
- We keep the ratio between the step size and the corresponding mesh element constant, i.e.

$$\frac{\Delta x_i}{x_i} = q = O(\varepsilon) \quad \Rightarrow \quad x_i = \frac{1}{1-q} x_{i-1}$$

- This mesh is quasi-linear in the sense of $\frac{\Delta x_{i-1}}{\Delta x_i} = 1 + O(q)$
- The Upwind and Lax-Wendroff schemes are then consistent on the progressive mesh
- Any accurate scheme for the PDE may be used (see e.g. T. Goudon, F. Lagoutière, L.M. Tine, 2013)

Models Calibration

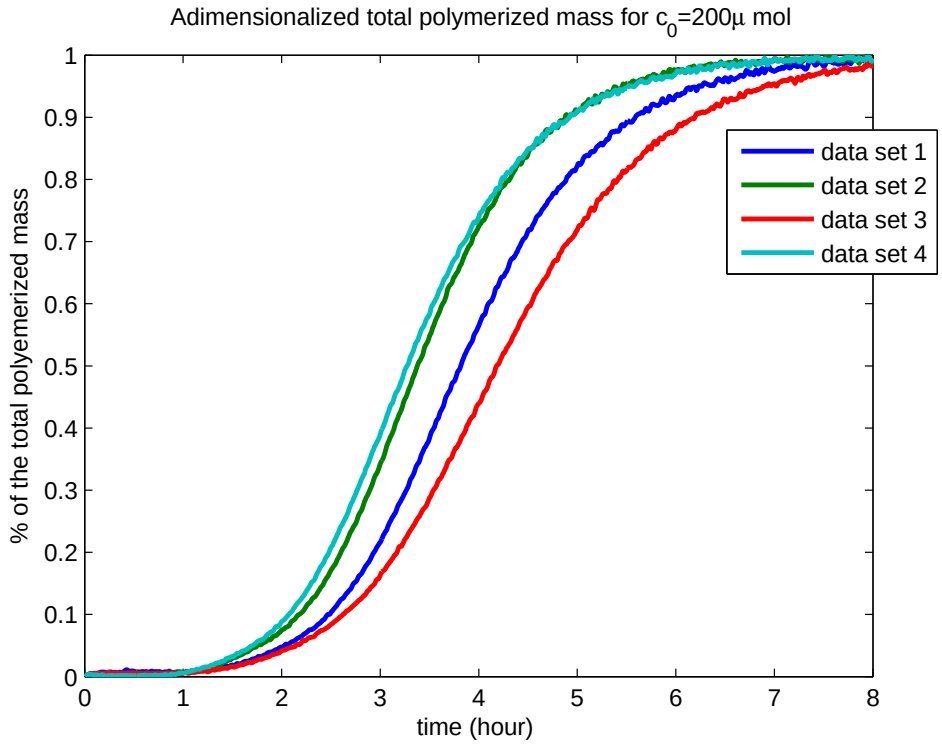
- Studies on PolyQ (Huntington's)

Models Calibration

- Studies on PolyQ (Huntington's)
- State Estimation on a toy model

PoyQ Aggregation

PolyQ: a domain in the protein HTT responsible for Huntington's Experiments done by H. Razaei and S. Prigent, INRA, (Virologie et Immunologie Moleculaires)



- Understand the key polymerization mechanisms
- How to select and calibrate the model?

PolyQ aggregation

First study (S. Prigent, A. Ballesta, MD et al., PLoS One, 2012)

Not an infectious disease: no coagulation, no fragmentation

$$\frac{dV^*}{dt} = -k_I^+ V^* + k_I^- V,$$

$$\frac{dV}{dt} = k_I^+ V^* - k_I^- V - \frac{i_0\, k_{on}^N\, V^{i_0+1} g(x_0)}{k_{off}^N + g(x_0) V} - V \int_{x_0}^{\infty} g\, n dx + \int_{x_0}^{\infty} g^- n dx,$$

$$\frac{\partial n}{\partial t} = -V \frac{\partial}{\partial x}(gn) + \frac{\partial}{\partial x}(g^- n)$$

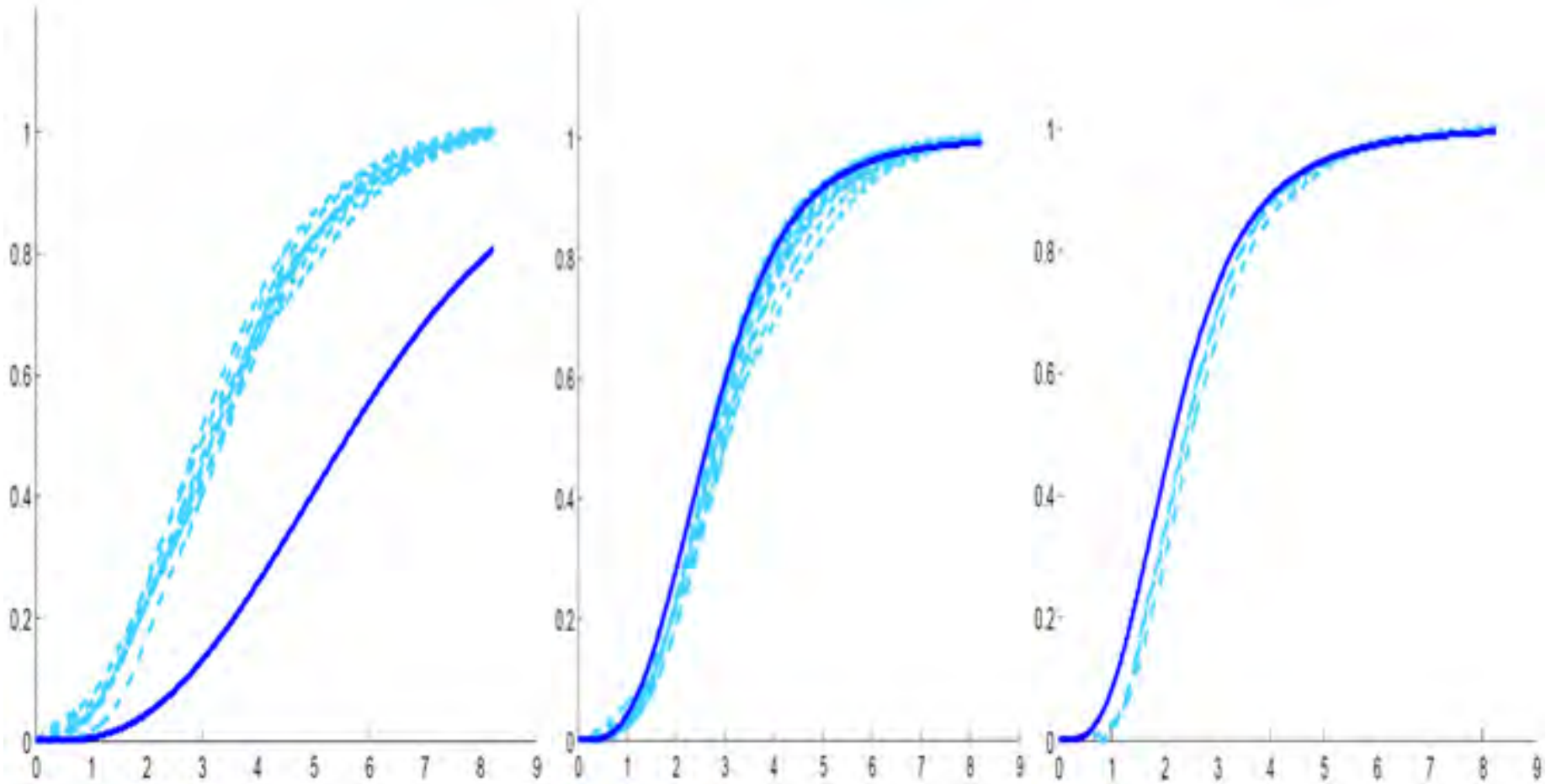
$$g(x_0) n(t, x_0) = g(x_0) \frac{k_{on}^N\, V^{i_0}}{k_{off}^N + g(x_0) V}.$$

An important question: what is the size of the nucleus i_0 ?

PolyQ aggregation

First study (S. Prigent, A. Ballesta et al., PLoS One, 2012)

$C_0 = 100\mu M$ $C_0 = 285\mu M$ $C_0 = 420\mu M$

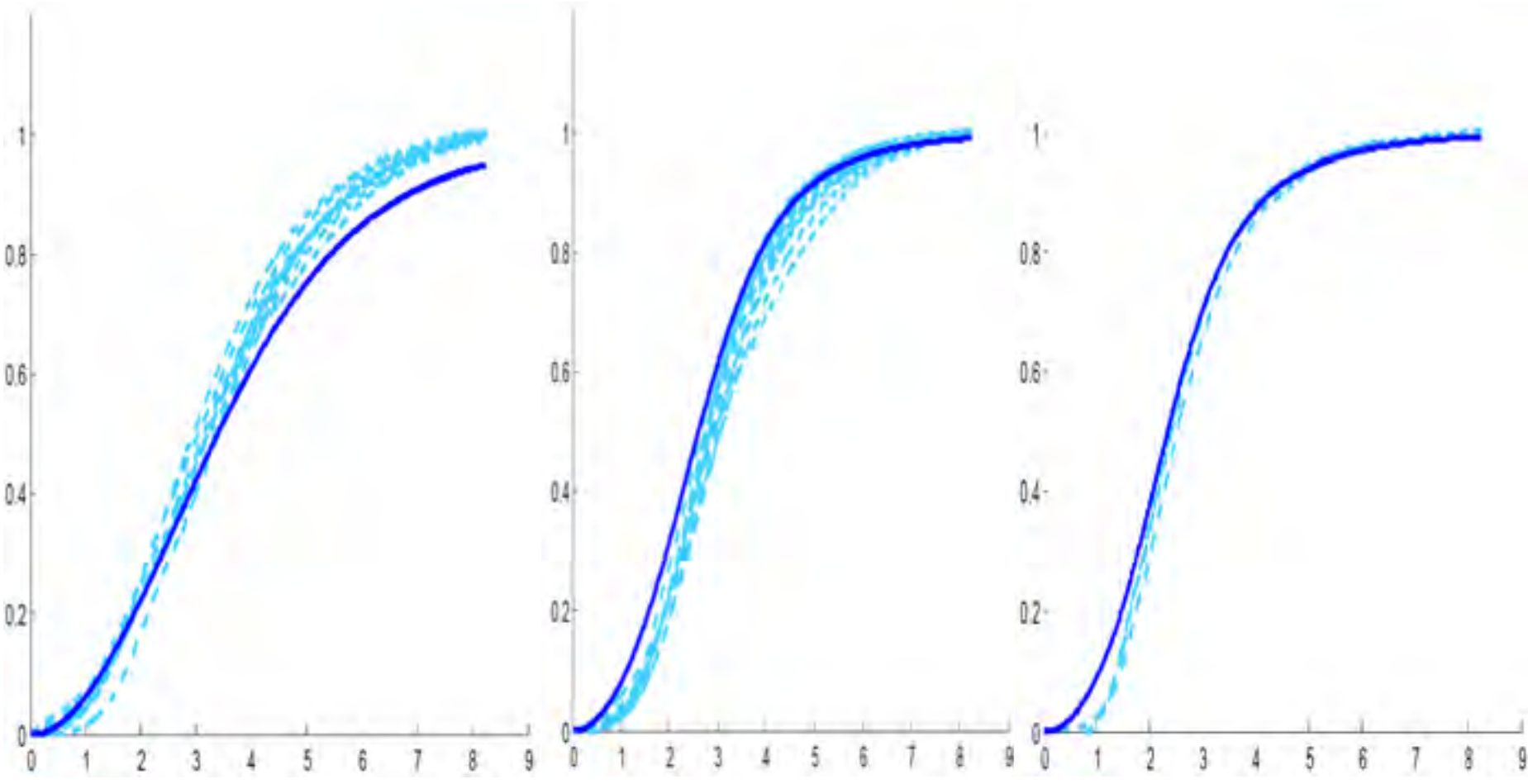


Best Fit for $i_0 = 3$: total error 40%

PolyQ aggregation

First study(S. Prigent, A. Ballesta et al., PLoS One, 2012)

$C_0 = 100\mu M$ $C_0 = 285\mu M$ $C_0 = 420\mu M$



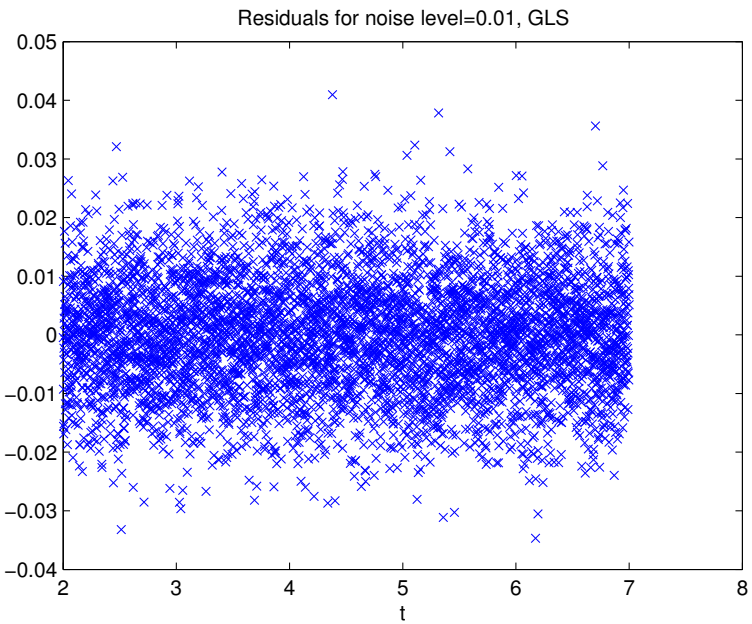
Best Fit for $i_0 = 1$: total error 10%

PolyQ aggregation

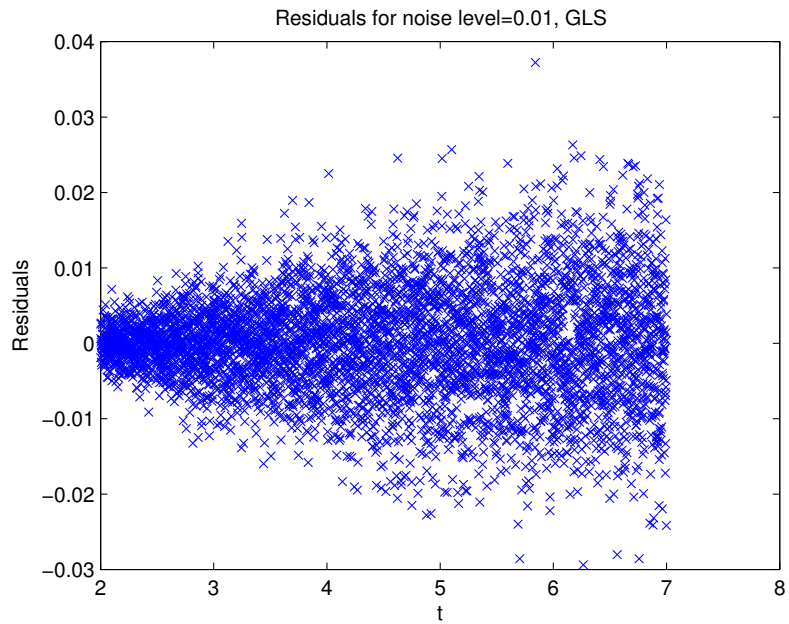
Systematic Parameters estimation study (H.T. Banks, MD, C. Kruse)

Ordinary Least Squares or Weighted Least Squares? Residual plots

$$r_i = \frac{data_i - M(t_i, \hat{\theta})}{M(t_i, \hat{\theta})^{\gamma}}, \quad \gamma \in [0, 1]$$



(a)

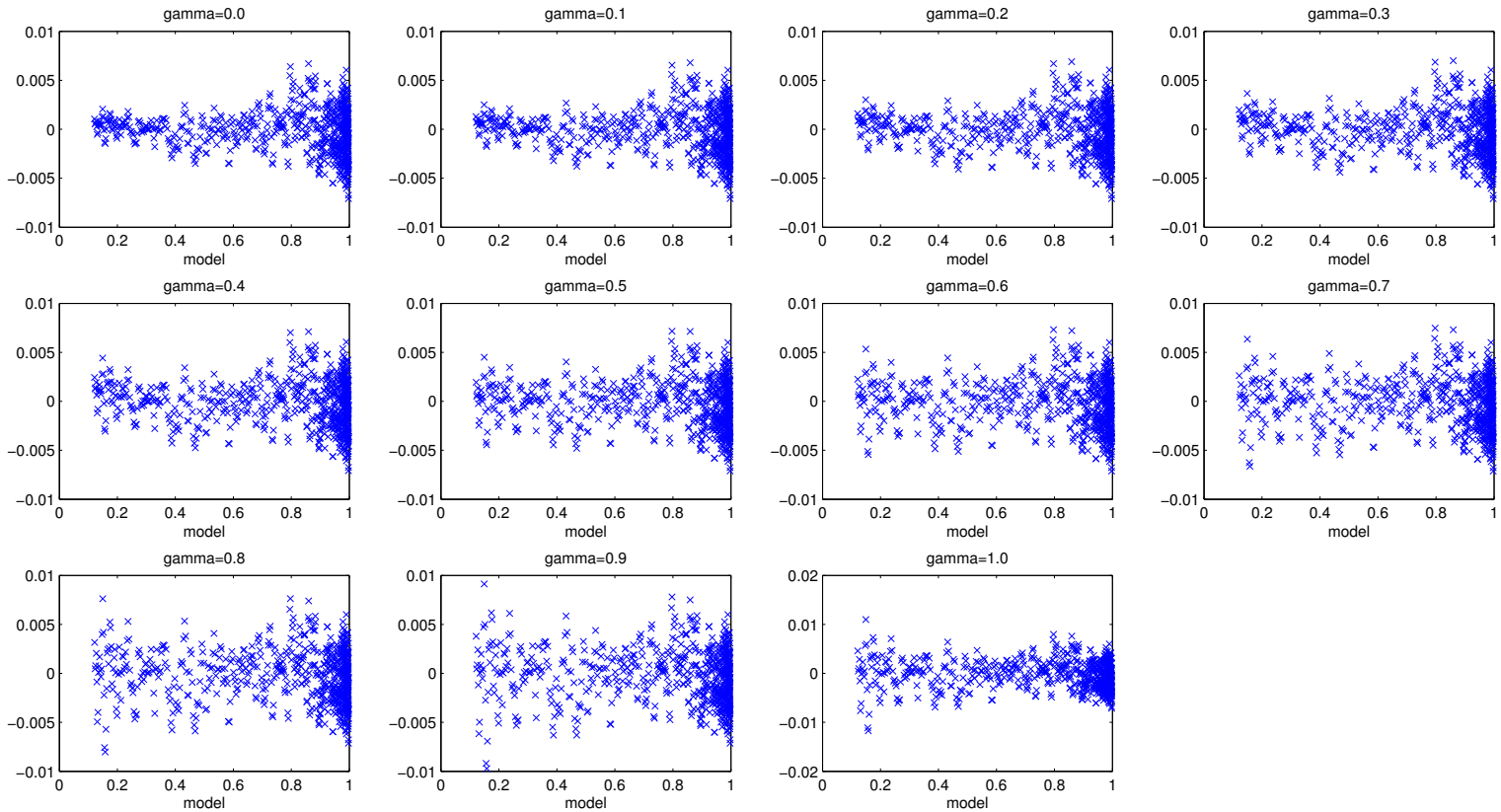


(b)

(a)Correct cost function ($\gamma = 1$); (b)Incorrect cost function ($\gamma = 0$)

Residual Plots - Weighted Least Squares with Varying γ

Inverse problem for datapoints ≥ 0.12 and $i_0 = 2$.



Standard Errors for Parameters Using WLS

Estimated parameters $\hat{\theta} = (k_I^+, ..., i_{max})$ (vector length $\kappa_\theta = 9$)
Use of the covariance matrix defined by

$$SE_k = \sqrt{\hat{\sigma}^2 (F)^{-1}},$$

for F defined from sensitivity analysis and σ^2 approximation of the variance.
For finite standard error using asymptotic theory, the 9×9 matrix $F = \chi^T(\hat{\theta}) W(\hat{\theta}) \chi(\hat{\theta})$ must be invertible.
Here: good fit of the curve and good residuals.

BUT: near linear dependence between certain rows of F, hence a huge condition number (10^{24}).

CONCLUSION: non uniqueness of the parameteters
At least for 1 curve at a time...

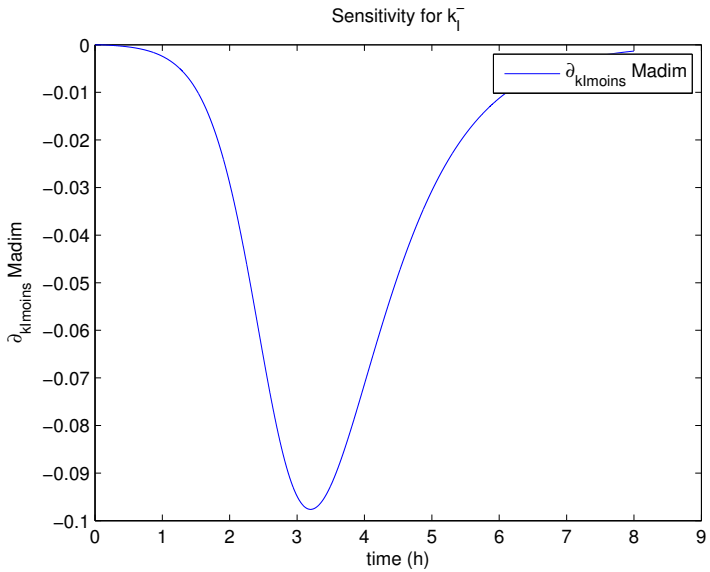
Sensitivity Analysis

Sensitivity matrices

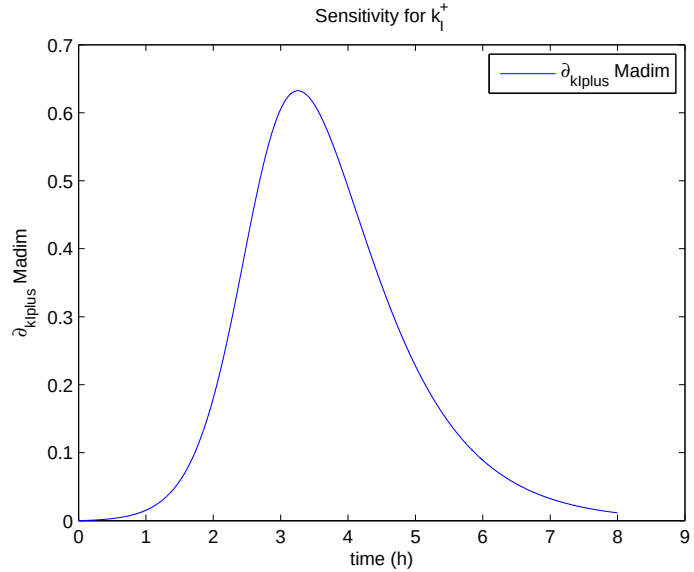
$$\chi = \frac{\partial M}{\partial \theta}$$

.

Mainly four parameters are sensitive: $k_l^+, k_l^-, k_{on}^N, k_{off}^N$.



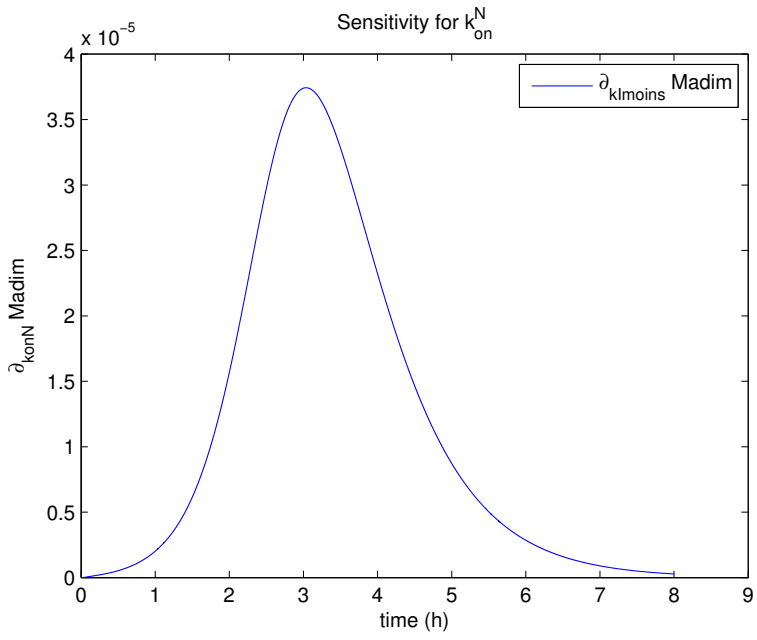
(a)



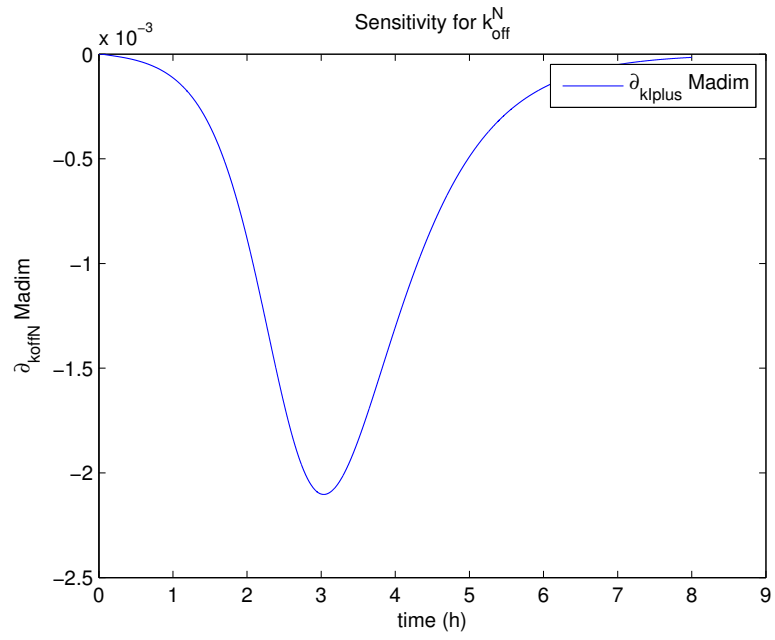
(b)

(a) Sensitivity w.r.t. k_l^- (b) Sensitivity w.r.t. k_l^+

Sensitivity Analysis (2/2)



(a)



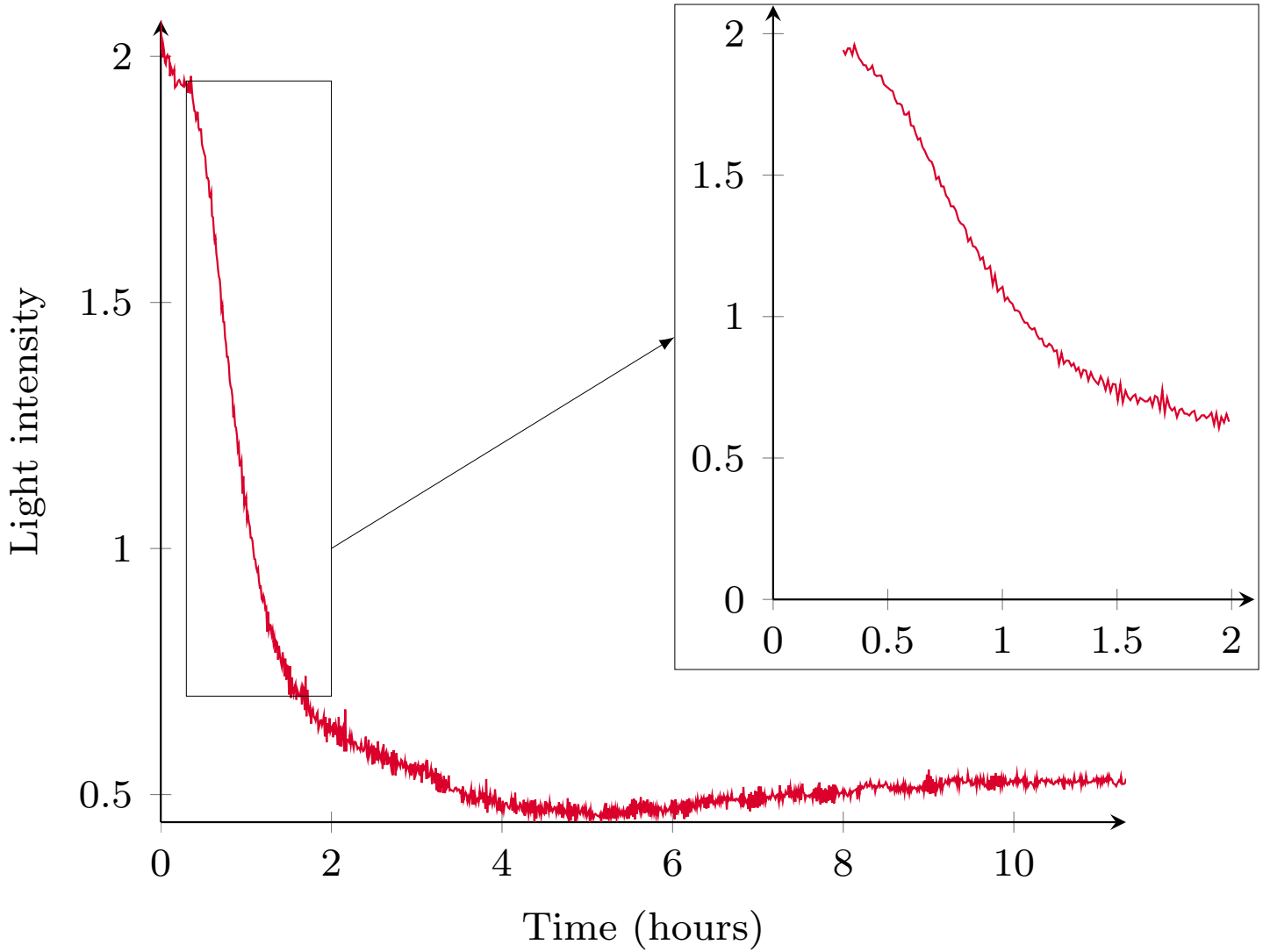
(b)

Figure : (a) Sensitivity w.r.t. k_{on}^N; (b) Sensitivity w.r.t. k_{off}^N

The sensitivities for the remaining parameters are in an order of magnitude of less than 10^{-6}.

State estimation: depolymerization case
(in progress with A. Armiento (PhD student) and P. Moireau)

`Real Experience` : Light intensity in SLS test



[2]Source: H. Rezaei's team, Lab. of Inra (Jouy-en-Josas, France)

Constant depolymerisation rate $k_{dep} \equiv b$

$$(LSs) \begin{cases} \frac{\partial}{\partial t} c(x,t) - \frac{\partial}{\partial x} bc(x,t) = 0, & (t,x) \in [0,T] \times [0,L], \\ \\ c(x,0) = c^{in}, \end{cases}$$

Analytic solution: $c(x,t) = c^{in}(x+bt)$.

`Observation =` $\mu_{\mathbf{n}}$

$$\mu_n[c^{in}](t) = \int_0^L x^n c(x,t)dx = \int_{bt}^L c^{in}(y)(y-bt)^n dy \qquad \text{for } n \in \mathbb{N}$$

`Actual Observation =` $\mu^{\mathbf{mes}}$

$$\mu^{mes} = \mu_n + \chi.$$

State estimation for a depolymerizing system

$$c^{in}(x) = \frac{1}{n!(-b)^{n+1}} \frac{d^{n+1}}{dt^{n+1}} \mu_n(\frac{x}{b}),$$

$$\mu^{mes} \quad \text{s.t.} \quad \|\mu^{mes} - \mu_n\|_{\mathcal{H}^{-\frac{1}{2}}([0,T])} \leq \varepsilon.$$

State estimation for a depolymerizing system

$$c^{in}(x) = \frac{1}{n!(-b)^{n+1}} \frac{d^{n+1}}{dt^{n+1}} \mu_n(\frac{x}{b}),$$

$$\mu^{mes} \quad \text{s.t.} \quad \|\mu^{mes} - \mu_n\|_{\mathcal{H}^{-\frac{1}{2}}([0,T])} \leq \varepsilon.$$

Regularization by filtering:

$$c_\varepsilon^{in} = \frac{1}{n!(-b)^{n+1}} \frac{d^{n+1}}{dt^{n+1}} \mu_n^\varepsilon(\frac{x}{b}).$$

State estimation for a depolymerizing system

$$c^{in}(x) = \frac{1}{n!(-b)^{n+1}} \frac{d^{n+1}}{dt^{n+1}} \mu_n(\frac{x}{b}),$$

$$\mu^{mes} \quad \text{s.t.} \quad \|\mu^{mes} - \mu_n\|_{\mathcal{H}^{-\frac{1}{2}}([0,T])} \leq \varepsilon.$$

Regularization by filtering:

$$c_\varepsilon^{in} = \frac{1}{n!(-b)^{n+1}} \frac{d^{n+1}}{dt^{n+1}} \mu_n^\varepsilon(\frac{x}{b}).$$

$$c_{\varepsilon,\alpha}^{in} = \rho_\alpha * c_\varepsilon^{in} = \rho_\alpha^{(n+1)} * \left(\frac{1}{n!(-b)^{n+1}} \mu_n^\varepsilon(\frac{x}{b}) \right).$$

Where $\rho_\alpha = \frac{1}{\alpha}\rho(\frac{x}{\alpha})$, $\rho \in \mathcal{C}_0^\infty(\mathbb{R}_+)$ s.t. $\mu_0[\rho] = 1$, $\mu_{1\leq k\leq m}[\rho] = 0$.

Proposition (Approximation Accuracy)

Given all the elements defined as before, we have

$$\|c_{\varepsilon,\alpha}^{in} - c^{in}\|_{\mathbb{L}^p} \leq C\left(\frac{\varepsilon}{\alpha^{n+0.5+1}} + \alpha^{m+1}\right) = F_\varepsilon(\alpha),$$

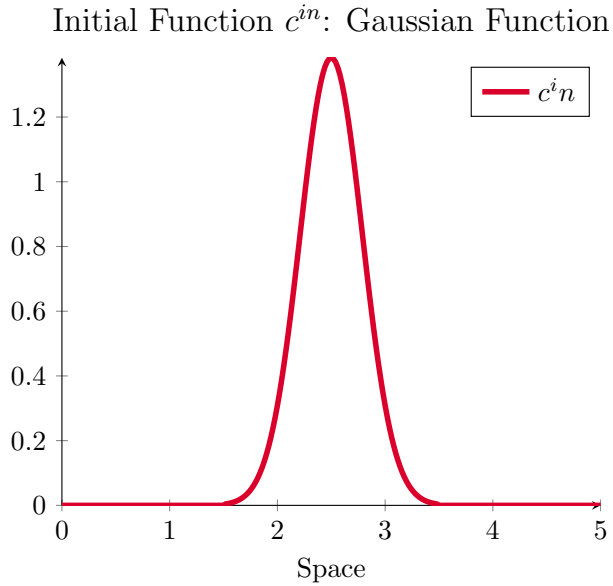
Furthermore, the optimal choice for the regularization parameter α *is*

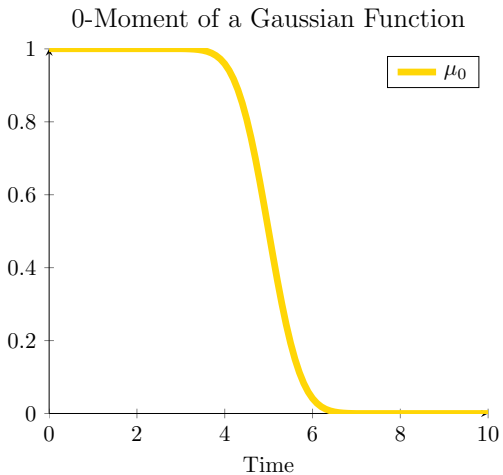
$$\alpha_{opt} = \varepsilon^{\frac{1}{n+m+2+s}}, \tag{1}$$

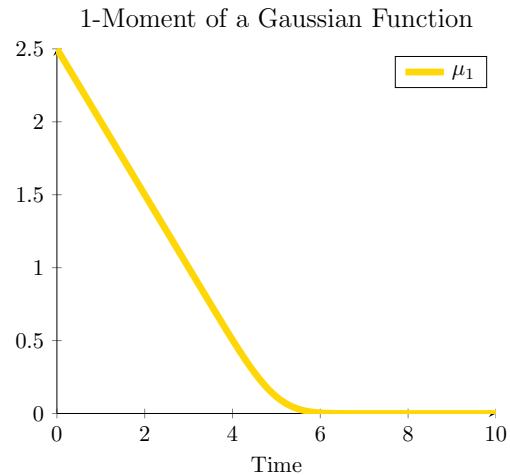
and, consequently, we have the following estimation

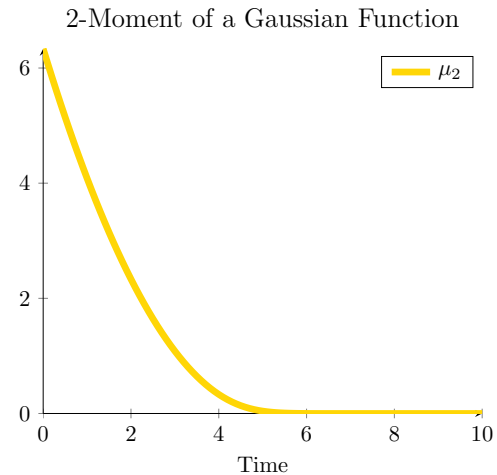
$$err_{opt} = \|c_{\varepsilon,\alpha_{opt}}^{in} - c^{in}\|_{\mathbb{L}^p} = O\left(\varepsilon^{\frac{m+1}{n+m+2+0.5}}\right). \tag{2}$$

Moments For a Gaussian Initial Condition

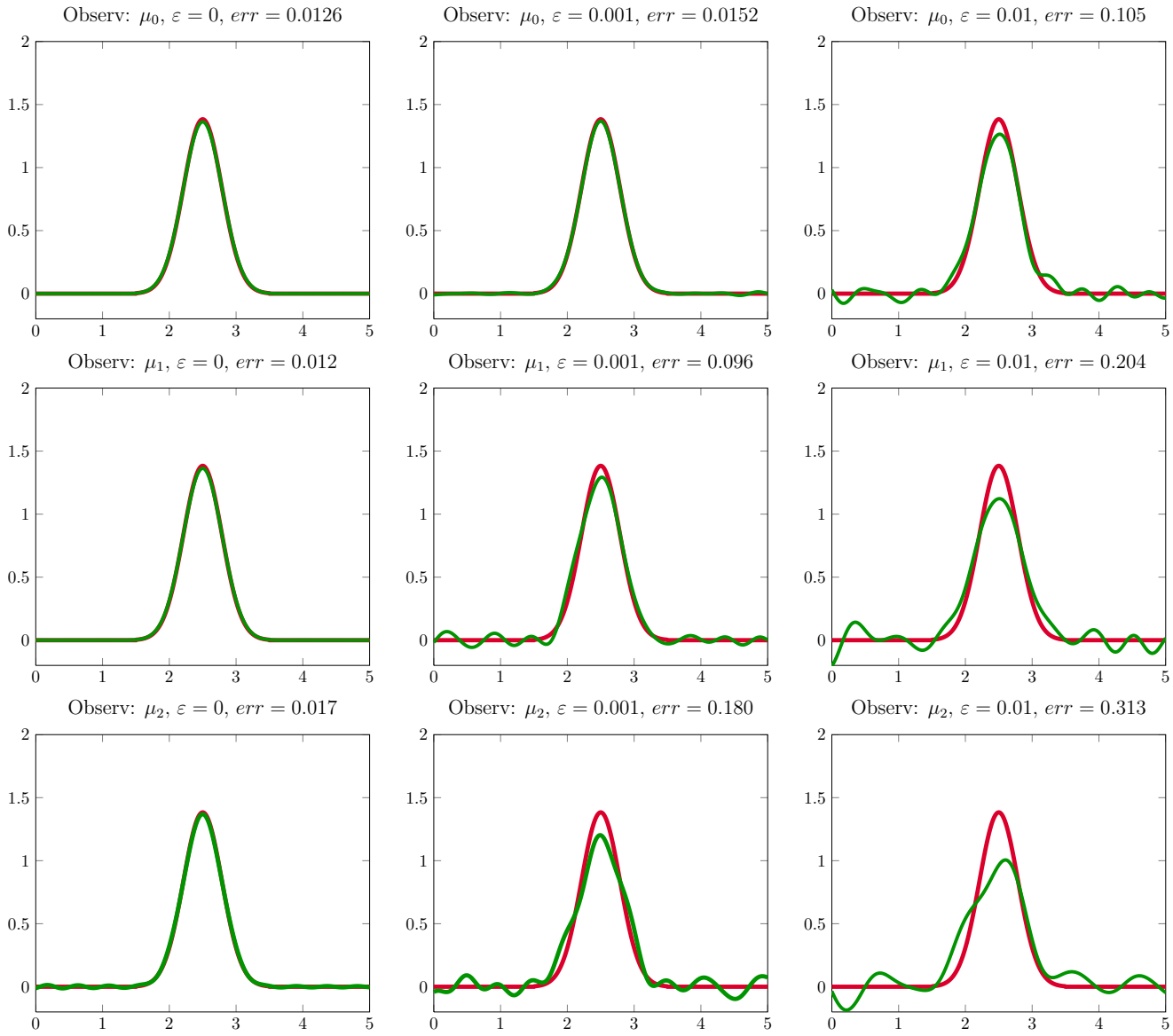








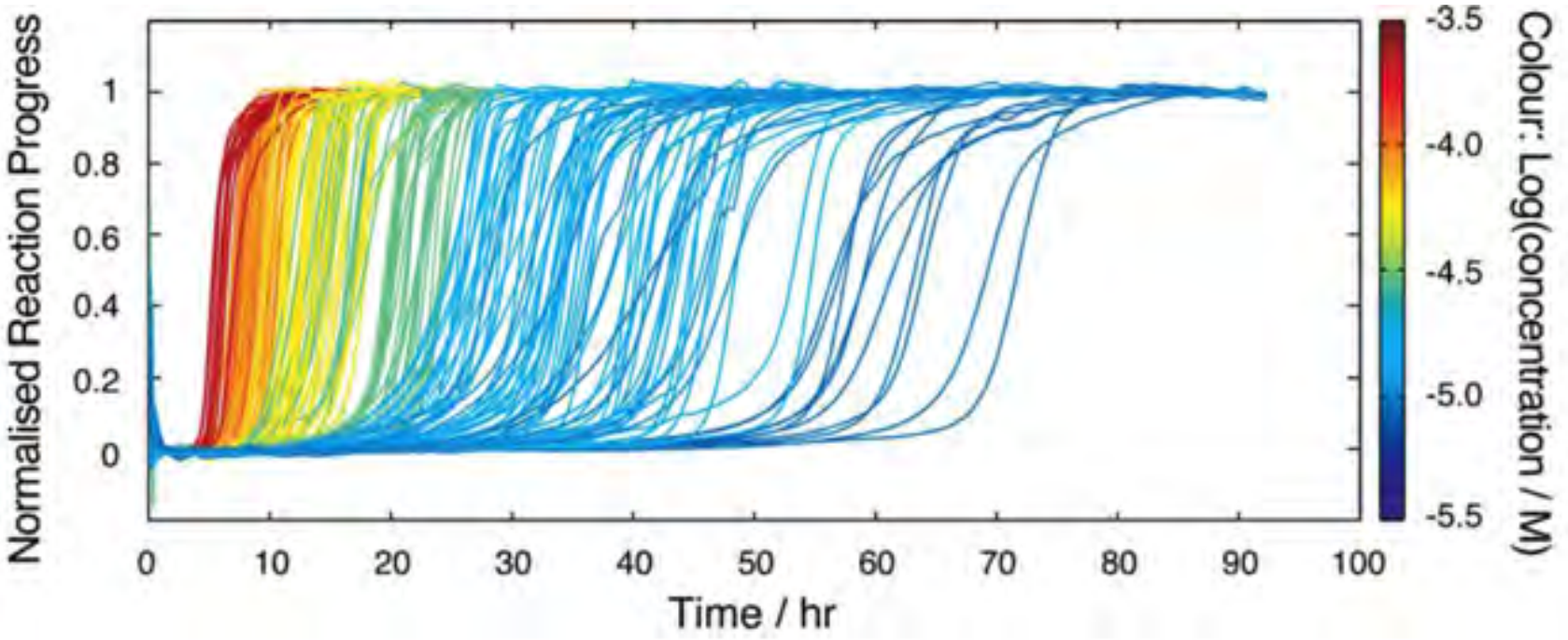
Simulations



Modelling variability among curves

Deterministic models: are they relevant?

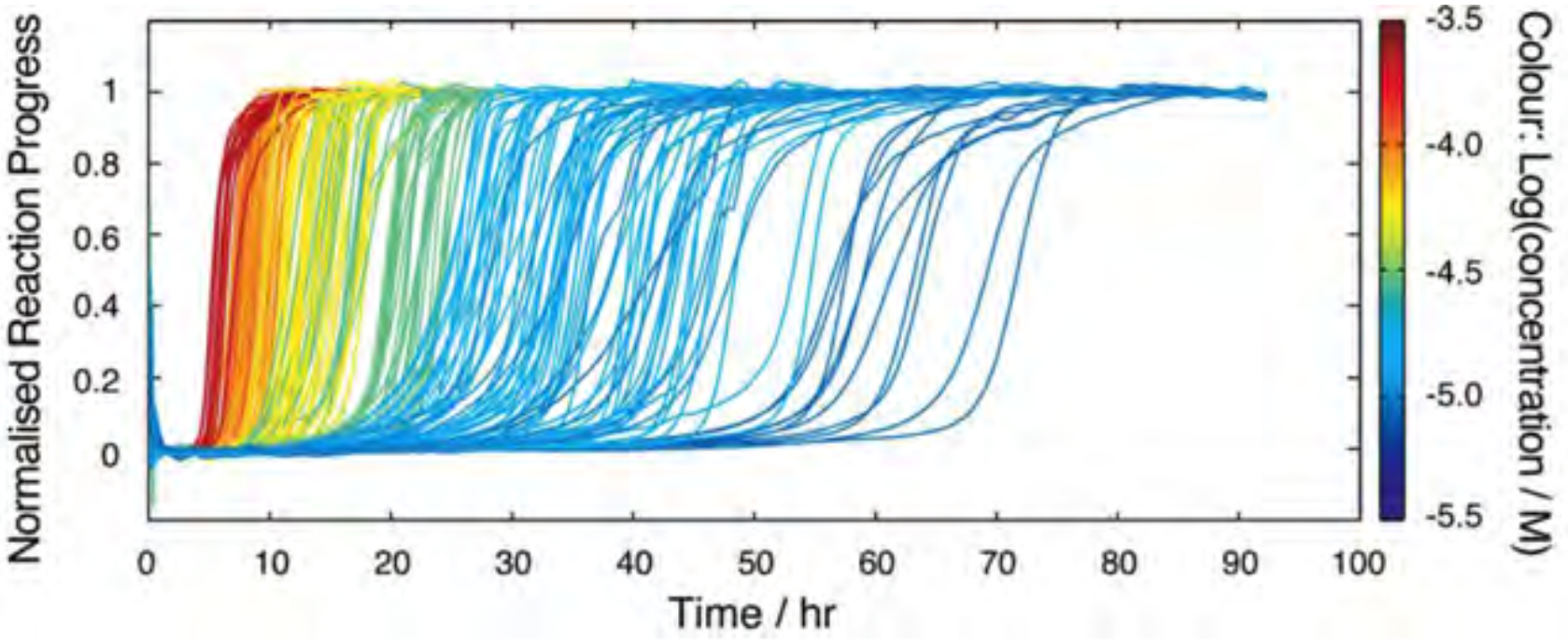
(work in progress with S. Eugene & P. Robert)



In vitro polymerization of $\beta 2m$. From Xue, Radford et al., PNAS, 2008

Deterministic models: are they relevant?

(work in progress with S. Eugene & P. Robert)



In vitro polymerization of $\beta 2m$. From Xue, Radford et al., PNAS, 2008
To take into account intrinsic variability:

- stochastic model for nucleation
- law of large number / variation of the lag time ?

Experimental Curves

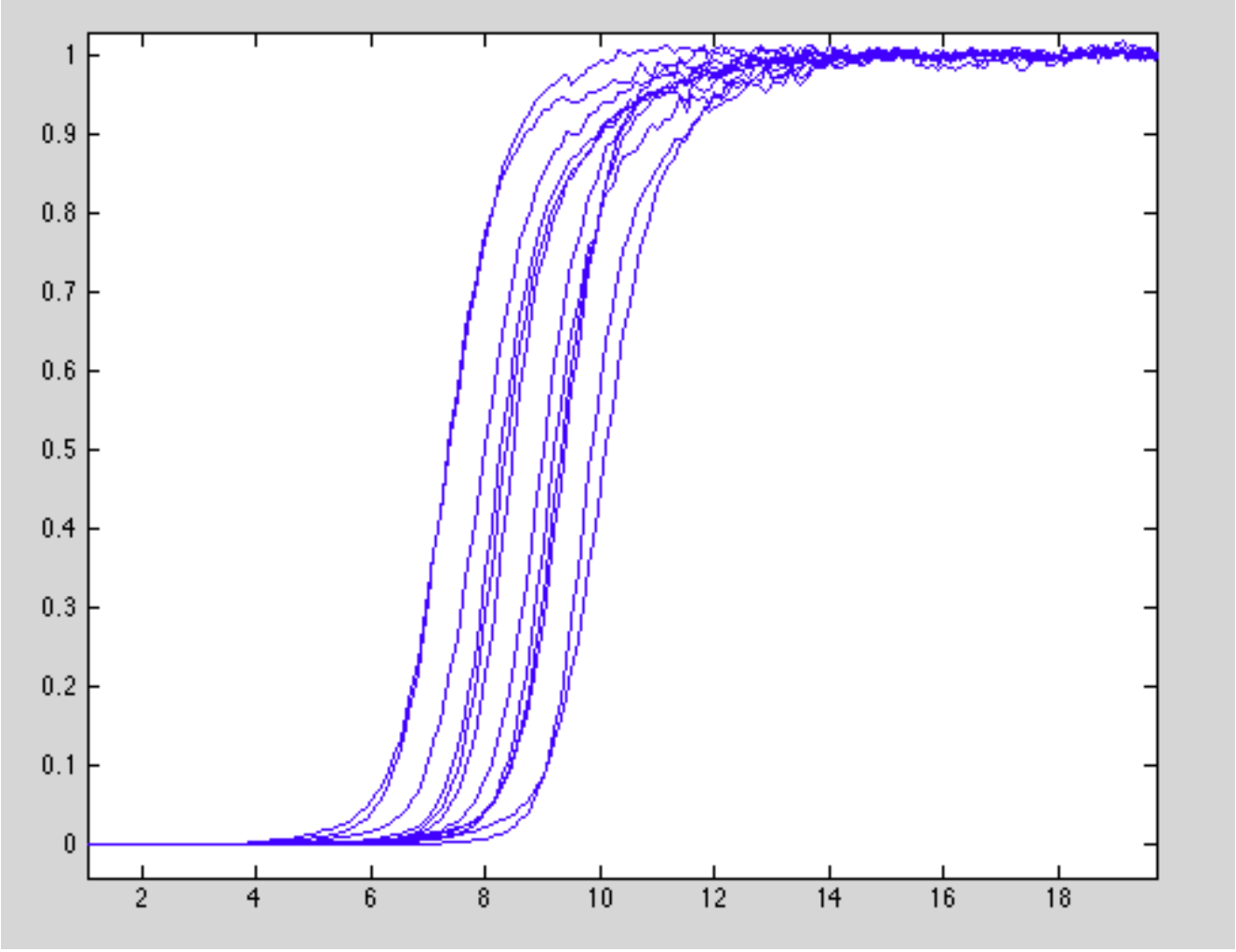


Figure : $m = 122\ \mu M$

Stochastic model (with S. Eugène and P. Robert)

- M monomers at $t = 0$
- large volume N
- $M/N = m$ remains constant (=initial concentration)
- $X(t)$: number of monomers at time t
- $\tilde{X}(t)$: number of polymers at time t

Stochastic model (with S. Eugène and P. Robert)

- M monomers at $t = 0$
- large volume N
- $M/N = m$ remains constant (=initial concentration)
- $X(t)$: number of monomers at time t
- $\tilde{X}(t)$: number of polymers at time t

$$\begin{cases} x + x & \xrightarrow{\alpha \cdot \left(\frac{X}{N}\right)^2} & 2\tilde{x} \\ x + \tilde{x} & \xrightarrow{\beta \cdot \frac{X}{N} \cdot \frac{\tilde{X}}{N}} & 2\tilde{x} \end{cases}$$

where $\alpha << \beta$. Stochastic equation, sum of Poisson processes:

$$\frac{dX}{dt} = -2 \sum_{i=1}^{X(X-1)/2} \mathcal{N}^i_{\alpha/N^2}(dt) - \sum_{i=1}^{X(M-X)} \mathcal{N}^i_{\beta/N^2}(dt)$$

First Order: Law of Large Numbers (with S. Eugène and P. Robert)

Proposition

The process $\overline{X}^N(t) = \frac{X(Nt)}{N}$ converges in distribution to $(x(t))_{t\geq 0}$, where $(x(t))_{t\geq 0}$ is the solution of :

$$\begin{cases} \dot{x} = -2\alpha x^2 - \beta x(m-x) \\ x(0) = m \end{cases}$$

Proof : Stochastic calculus methods

Second Order: Central Limit Theorem (with S. Eugène and P. Robert)

Proposition

The process $\left(\frac{X(Nt)-Nx(t)}{\sqrt{N}}\right)_{t\geq 0}$ converges in distribution to $(U(t))_{t\geq 0}$, where $U(t)$ is solution of the SDE :

$$dU_t = \frac{\beta m\sqrt{\alpha}e^{\beta mt/2}}{\alpha e^{\beta mt}+\beta-\alpha}dW_t + \beta m^2\left[\frac{\beta-\alpha-\alpha e^{\beta mt}}{\beta-\alpha+\alpha e^{\beta mt}}\right]U_t dt$$

Proof : Stochastic calculus methods

Second Order: Central Limit Theorem (with S. Eugène and P. Robert)

Proposition

The process $\left(\frac{X(Nt)-Nx(t)}{\sqrt{N}}\right)_{t\geq 0}$ converges in distribution to $(U(t))_{t\geq 0}$, where $U(t)$ is solution of the SDE :

$$dU_t = \frac{\beta m\sqrt{\alpha}e^{\beta mt/2}}{\alpha e^{\beta mt}+\beta-\alpha}dW_t + \beta m^2\left[\frac{\beta-\alpha-\alpha e^{\beta mt}}{\beta-\alpha+\alpha e^{\beta mt}}\right]U_t dt$$

Proof : Stochastic calculus methods

Conclusion

$$\frac{X(Nt)}{N} \underset{N\to\infty}{\sim} x(t) + \frac{1}{\sqrt{N}}U(t)$$

Application: Lag Time

Definition

We define the lag time by the following random variable :

$$T_N^{\delta} = \inf\{t > 0, X(t) \leq \delta N\}$$

Application: Lag Time

Definition

We define the lag time by the following random variable :

$$T_N^\delta = \inf\{t > 0, X(t) \leq \delta N\}$$

Corollary

As N goes to infinity :

$$\frac{T_\delta^N}{N} \underset{N\to\infty}{\sim} t_\delta + \frac{1}{\sqrt{N}} \frac{U(t_\delta)}{-\dot{x}(t_\delta)}$$

where $t_\delta = x^{-1}(\delta)$.

Application: Lag Time

$$\frac{T_\delta^N}{N} \underset{N\to\infty}{\sim} t_\delta + \frac{1}{\sqrt{N}} \frac{U(t_\delta)}{-\dot{x}(t_\delta)}$$

Proposition

The **random** variable $\frac{U(t_\delta)}{-\dot{x}(t_\delta)}$ is almost independent of delta.

Application: Lag Time

$$\frac{T_\delta^N}{N} \underset{N\to\infty}{\sim} t_\delta + \frac{1}{\sqrt{N}} \frac{U(t_\delta)}{-\dot{x}(t_\delta)}$$

Proposition

The **random** variable $\frac{U(t_\delta)}{-\dot{x}(t_\delta)}$ is almost independent of delta.

Conclusion

For a given initial concentration of monomers, the polymerisation curves are translations of the mean $x(t)$.

Experimental Curves

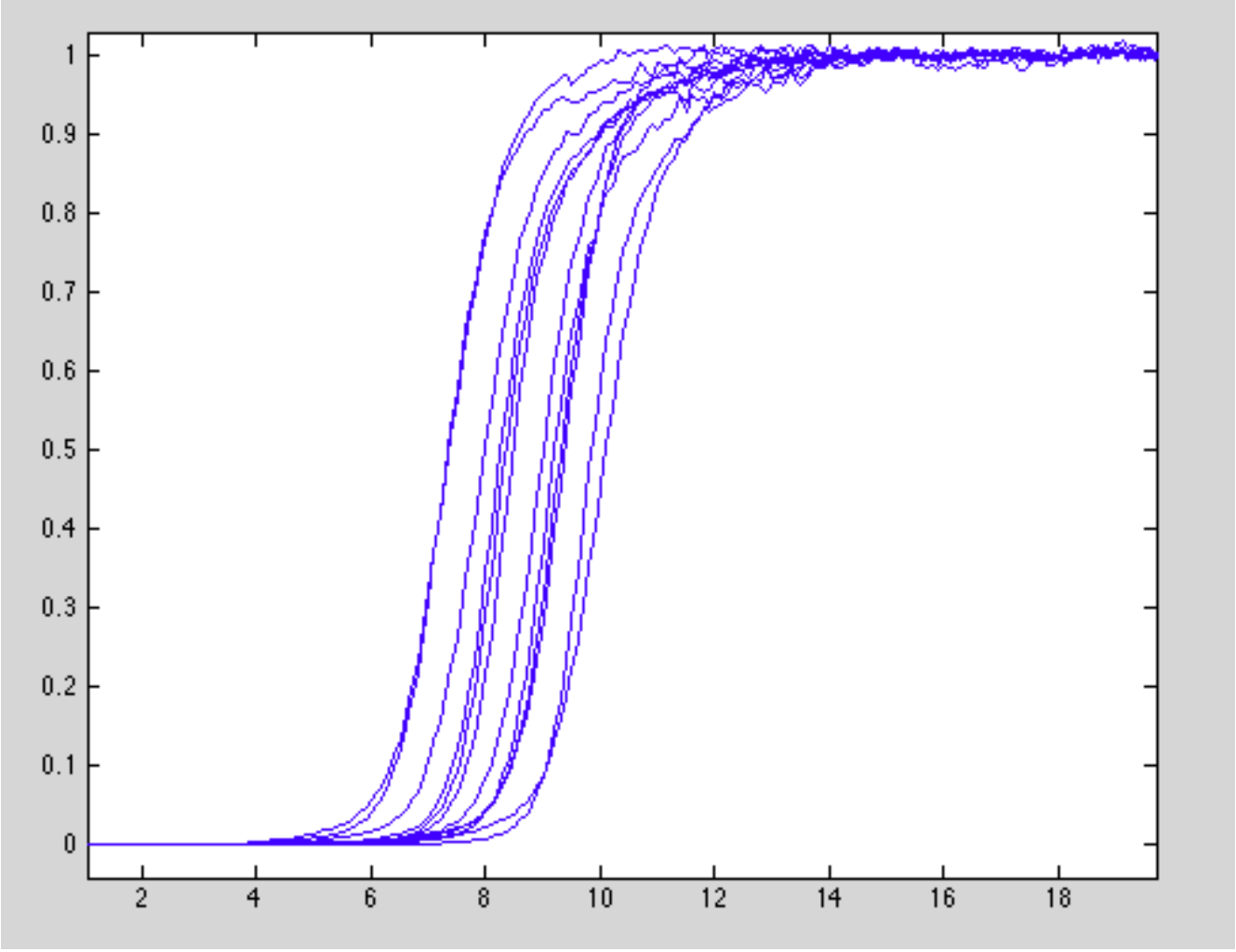


Figure : $m = 122\ \mu M$

Translated Experimental Curves

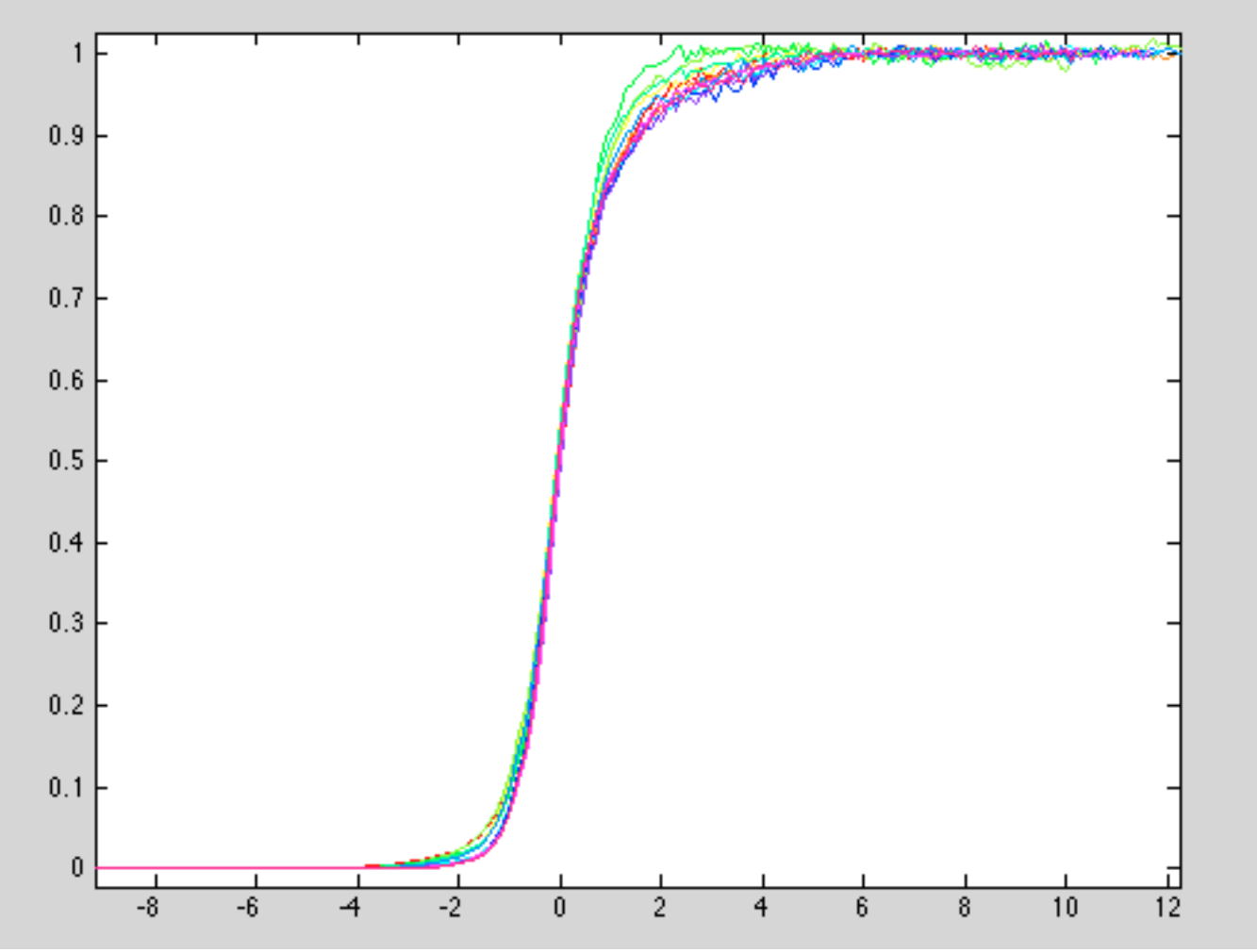


Figure : $m = 122\ \mu M$

Estimation of the parameters

$$\begin{cases} x + x \xrightarrow{\alpha \cdot \left(\frac{X}{N}\right)^2} 2\tilde{x} \\ x + \tilde{x} \xrightarrow{\beta \cdot \frac{X}{N} \cdot \frac{\tilde{X}}{N}} 2\tilde{x} \end{cases}$$

$\rightarrow$ Only **two** parameters : α and β.

- α controls the beginning of the reaction
- β controls the take-off

Estimation of α

Numerical results

m (μM)	$\alpha(10^{-6}.h^{-1}.\mu M^{-1})$
16.7	1.91×10^{-5}
30.5	1.31×10^{-7}
43.7	0.0056
61	0.0279
84.1	0.0014
102.2	0.0151
122	0.0035
142.1	0.0219
243.5	0.3268

$\Rightarrow$ **Non robust estimation**: due to variance?

Estimation of β

Numerical results

m (μM)	$\beta(10^{-2}.h^{-1}.\mu M^{-1})$
16.7	4.53
30.5	4.54
43.7	2.61
61	2.82
84.1	2.32
102.2	1.78
122	1.59
48.5	2.33
142.1	1.41
243.5	0.77

$\Rightarrow$ **Robust estimation**: close β for different initial concentrations.

Estimation of the variance

Observe that for $\frac{\beta}{\alpha} >> 1$:

$$\begin{aligned}
\mathrm{Var}\left(\frac{T_\delta^N}{N}\right) &\approx \frac{1}{Nm^3\beta^2}\left(\frac{\delta}{1-\delta}\right)\left\{2\frac{1-\delta}{\delta}\ln\left(\frac{\beta}{\alpha}\left(\frac{\delta}{1-\delta}\right)\right)+\frac{\beta}{\alpha}\frac{1-\delta}{\delta}\right\} \\
&\approx \frac{1}{N\alpha\beta m^3} \quad \text{(independent of } \delta\text{)} \\
&\approx \frac{1}{M\alpha\beta m^2}
\end{aligned}$$

For the data we have:

$$\mathrm{Var}\left(\frac{T_\delta^N}{N}\right) \approx 10^{-9} hour...$$

Modified Stochastic Model

$$\begin{cases} y & \underset{k_I^-}{\overset{k_I^+}{\rightleftharpoons}} & x, \\ x + x & \xrightarrow{\tilde{\alpha}\cdot\left(\frac{X}{N}\right)^2} & 2\tilde{x} \\ x + \tilde{x} & \xrightarrow{\tilde{\beta}\cdot\frac{X}{N}\cdot\frac{\tilde{X}}{N}} & 2\tilde{x} \end{cases}$$

If instantaneous equilibrium between x and y : with $r = \frac{k_I^+}{k_I^-} \ll 1$:

$$x(t) \approx r y(t), \qquad \forall\, t,$$

and we calculate a new variance for the lag time

$$\mathrm{Var}\left(\frac{T_\delta^N}{N}\right) \approx \frac{1}{M r \alpha \beta m^2},$$

with the previously estimated $\alpha = \tilde{\alpha}$ and $\beta = \frac{1}{r}\tilde{\beta}$.

Experimental Data and Simulations - $r \approx 10^{-9}$

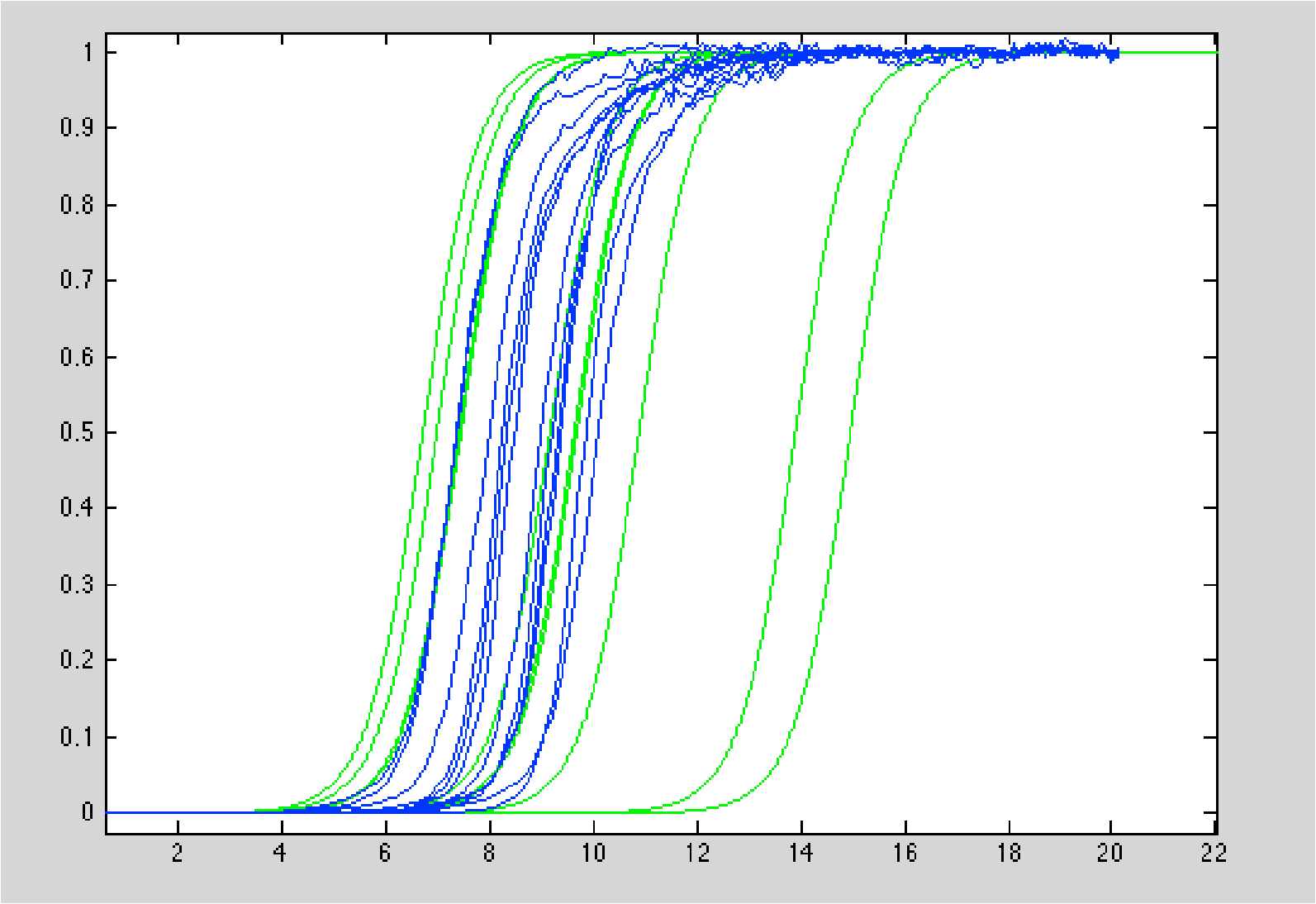


Figure : $m = 122\ \mu M$

Conclusion

- Stochastic model:
  - predicts an intrisic variability of any possible intensity with only 3 parameters
  - No need of experimental noise
  - Robust estimation of β
  - Use the variability to calibrate the model!
  - Ongoing work :
    extend stochastic model to include nucleation and fragmentation
- Aggregation models need to adapt to each situation
- Key importance of size distributions
- Inverse problems, state estimation: promising first results, still largely open

in collaboration with

INRA team of biologists in Jouy-en-Josas (VIM/MAP2)

H. Rezaei, J. Torrent, D. Martin

INRIA/LJLL MAMBA team

H.W. Haffaf (PhD), C. Kruse (post-doc),
S. Prigent (post-doc, **former at INRA**)

Other INRIA teams

P. Robert (RAP), S. Eugene (PhD, RAP & MAMBA)
P. Moireau (M3DISIM), A. Armiento (PhD, M3DISIM & MAMBA)

other collaborators

H.T. Banks (Raleigh, USA), M. Escobedo (BECAM, Spain), W.F. Xue (Canterbury)

References


Doumic, M. and Gabriel, P. (2010) *Eigenelements of a General Aggregation-Fragmentation Model*. Math. Models Methods Appl. Sci. 20(5), 757–783.


Doumic, M., Goudon, T., Lepoutre, T., *Scaling Limit of a Discrete Prion Dynamics Model*, Comm. Math. Sc., vol.7 Issue 4 (Dec. 2009), pp. 839–865.


Calvez, V. and Doumic, M. and Gabriel, P. (2010) *Self-similarity in a General Aggregation-Fragmentation Problem ; Application to Fitness Analysis*, J. de Math. Pur. et Appl., 2012.


Doumic, M., Hoffmann, M., Reynaud-Bouret, P. and Rivoirard, V. (2011) Nonparametric estimation of the division rate of a size-structured population. SIAM J. Numer. Anal., 2012.


Doumic, M., Perthame, B. and Zubelli, J. (2009) *Numerical Solution of an Inverse Problem in Size-Structured Population Dynamics*. Inverse Problems, 25, 25pp.


Doumic, M. and Tine, L.M., *Estimating the Division Rate of the Growth-Fragmentation Equation*, J. of Math. Biol., 2012, accepted.


Prigent, S. Ballesta, A. et al., *An Efficient Kinetic Model for Assemblies of Amyloid Fibrils and Its Application to Polyglutamine Aggregation*, Plos One, 2012.


Doumic, M., Hoffmann, M., Krell, N. and Robert, L., *Statistical Estimation of a Growth-Fragmentation Model Observed on a Genealogical Tree*, submitted.

2. Regularity of the map

generalization of a result of B. Perthame, J.P. Zubelli, Inv. Prob., 2007

Definition (Definition from MD, L.M. Tine, JMB, 2012)

Let g, k satisfying the assumptions of the "eigenvalue theorem". For a constant $b \geq 0$ and functions $f_0 \in L^1_0$, $f_\infty \xrightarrow[x\to+\infty]{} \infty$.

We define $\mathcal{D}(b, f_0, f_\infty) :=$

$$\left\{ B \in L^\infty_{\mathrm{loc}}(\mathbb{R}^*_+) \cap \mathbb{P}, \quad \mathrm{supp}(B) = [\tilde{b} \leq b, +\infty), \quad \frac{B}{g} \leq f_0, \quad \frac{xB}{g} \geq f_\infty \right\}.$$

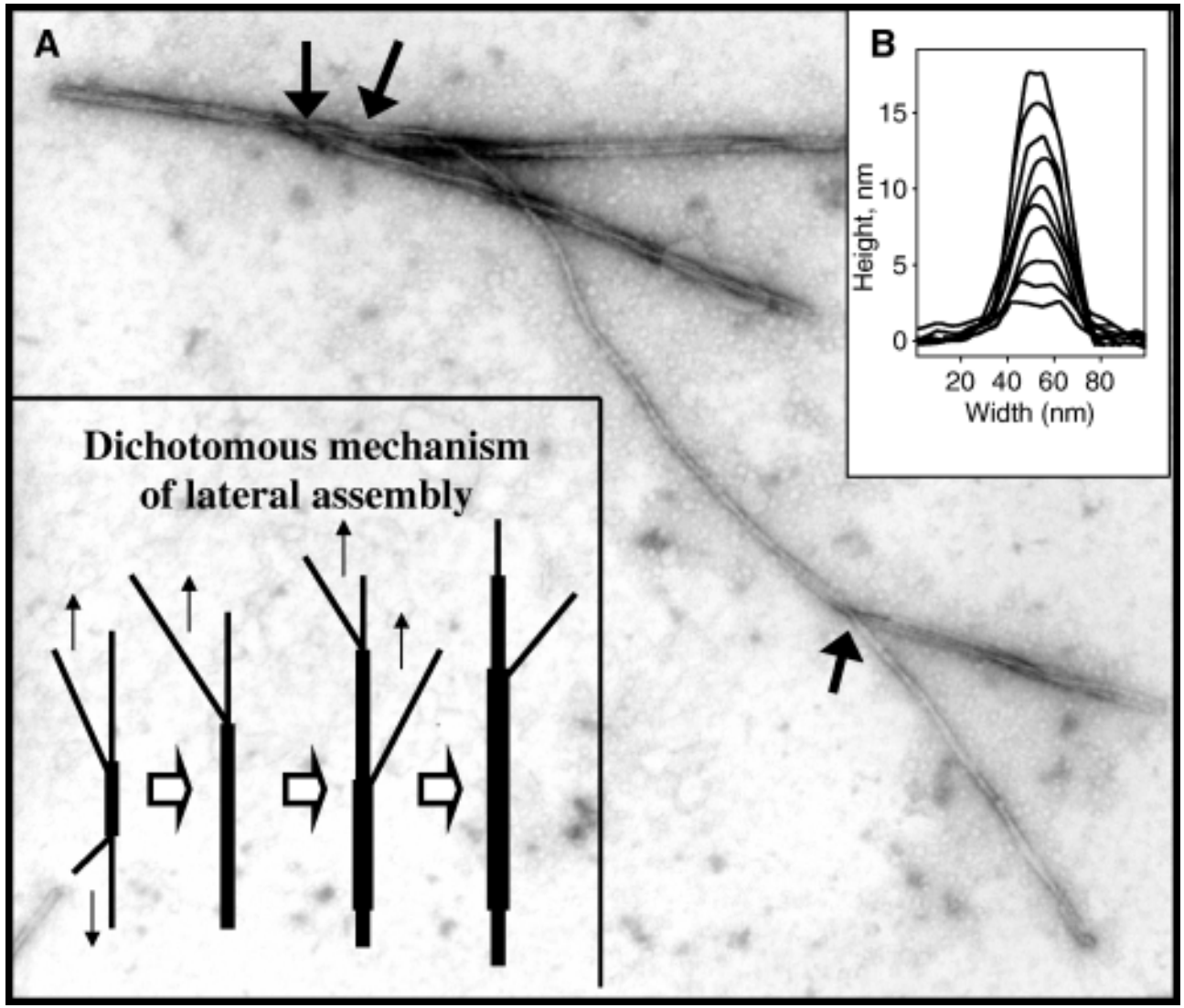
Theorem (from M.D., L.M. Tine, JMB, 2012)

Let parameters g and k satisfy the assumptions of this def. Then

i) $\Gamma : B \longmapsto (\lambda_0, N)$ *is injective and continuous under the* $L^\infty-$*weak-*topology for* B *from any set* $\mathcal{D}(b, f_0, f_\infty)$ *to* $\mathbb{R}^*_+ \times L^1(\mathbb{R}_+) \cap L^\infty(\mathbb{R}_+)$.

ii) *Let* $\frac{1}{g} \in L^2_0$. *Then* Γ *is locally* C^1 *under the strong topology of* $L^2(\mathbb{R}_+) \cap \mathcal{D}(b, f_0, f_\infty)$.

Protein polymerization: measures



Extensions of the model

Variability:

$$\frac{\partial}{\partial t} n(t,x,v) + \frac{\partial}{\partial x}\left(vxn(t,x,v)\right) =$$

$$-B(x)n(t,x,v) + 2\int_x^\infty \int_0^\infty B(y)k(y,x)\rho(v',v)n(t,y,v')dy,dv'$$

with $\int_0^\infty \rho(v',v)dv = 1$

Extensions of the model

Variability:

$$\frac{\partial}{\partial t} n(t,x,v) + \frac{\partial}{\partial x}\left(vxn(t,x,v)\right) =$$

$$-B(x)n(t,x,v) + 2\int_x^\infty \int_0^\infty B(y)k(y,x)\rho(v',v)n(t,y,v')dy,dv'$$

with $\int_0^\infty \rho(v',v)dv = 1$

Age + variability:

$$\frac{\partial}{\partial t} n(t,a,x,v) + \frac{\partial}{\partial x}\left(vxn(t,a,x,v)\right) = -B(a,x)n(t,a,x,v),$$
$$n(t,a=0,x,v) = 2\int_x^\infty \int_0^\infty B(a,y)k(y,x)\rho(v',v)n(t,a,y,v')dy\,dv'\,da$$

(related (maturity) models: Lebowitz, Rubinow, 1977 - Rotenberg, 1983 - Mischler, Perthame, Ryzhik, 2002,...)

Numerical Schemes

(H.T. Banks, MD, C. Kruse, submitted)

Example on a simple nucleation+polymerization case:

$$\frac{dV}{dt} = -k_l^+ V + k_l^- V^*,$$

$$V^* = c_0 - V - \sum_{i=i_0}^{N_0} i c_i - \int_{N_0}^{\infty} x c^\epsilon \, dx,$$

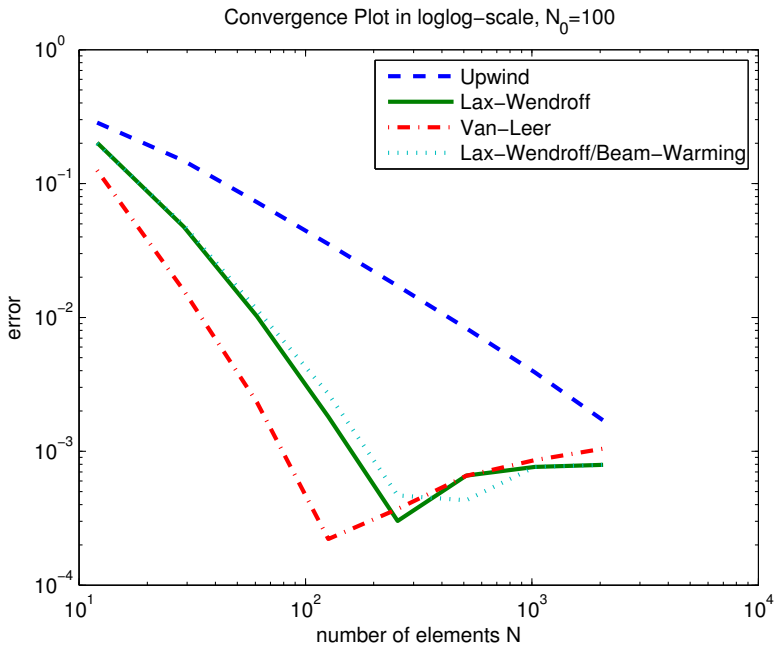
$$\frac{dc_{i_0}}{dt} = k_{on}^N (V^*)^{i_0} - k_{off}^N c_{i_0} - k_{on}^{i_0} c_{i_0} V^*,$$

$$\frac{dc_i}{dt} = V^* (k_{on}^{i-1} c_{i-1} - k_{on}^i c_i), \qquad i \leq N_0$$

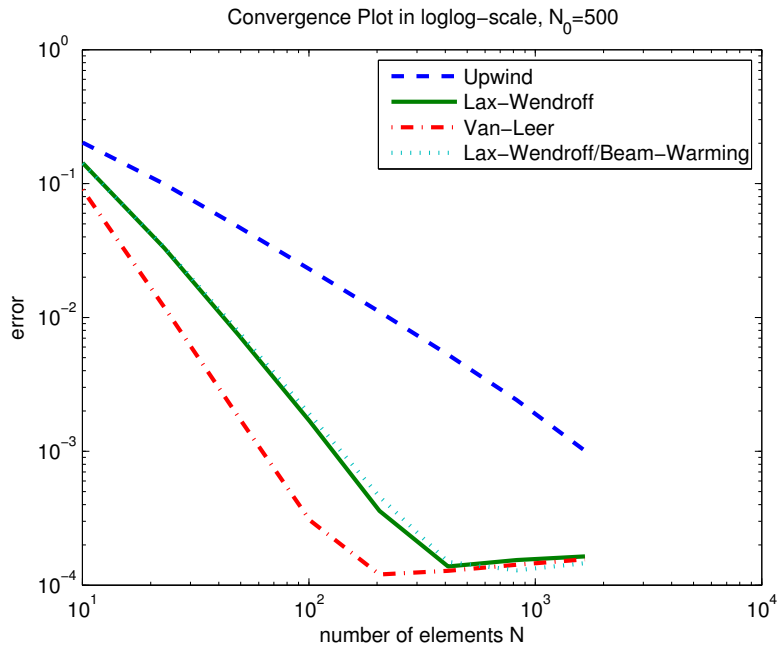
$$\partial_t c^\epsilon(x,t) = -V^* \partial_x (k_{on} c^\epsilon(x,t)), \qquad x \geq N_0$$

Numerical Schemes

(H.T. Banks, MD, C. Kruse, submitted)



(a)



(b)

(a) Convergence plot for $N_0 = 100$; (b) for $N_0 = 500$

Standard Errors for Parameters Using WLS

Estimated parameters $\hat{\theta} = (k_I^+, ..., i_{max})$ (vector length $\kappa_\theta = 9$)
Covariance matrix defined by

$$SE_k = \sqrt{\Sigma_{kk}(\hat{\theta})}, \qquad k = 1, ..., 9,$$

where

$$\Sigma(\hat{\theta}) = \hat{\sigma}^2(\chi^T(\hat{\theta})W(\hat{\theta})\chi(\hat{\theta}))^{-1}.$$

Here χ is the sensitivity matrix of size $n \times \kappa_\theta$ (n number of data points and κ_θ number of estimated parameters) and W defined by

$$W^{-1}(\hat{\theta}) = \text{diag}(M(t_1; \hat{\theta})^{2\gamma}, \ldots, M(t_n; \hat{\theta})^{2\gamma}).$$

Approximation of the variance

$$\sigma^2 \approx \hat{\sigma}(\hat{\theta})^2 = \frac{1}{n - \kappa_\theta} \sum_{i=1}^{n} \frac{1}{M(t_i; \hat{\theta})^{2\gamma}} (M(t_i, \hat{\theta}) - data_i)^2.$$

Standard Errors for Parameters Using WLS

To obtain a finite standard error using asymptotic theory, the 9×9 matrix $F = \chi^T(\hat{\theta})W(\hat{\theta})\chi(\hat{\theta})$ must be invertible.
Aboveseen problem: good fit of the curve and good residuals.

BUT:

- Condition number of $F = \chi^T(\hat{\theta})W(\hat{\theta})\chi(\hat{\theta})$ is $\kappa = 10^{24}$.
- near linear dependence between certain rows of F, hence the large condition number.

CONCLUSION:

- non uniqueness of the parameters
- At least for 1 curve at a time...

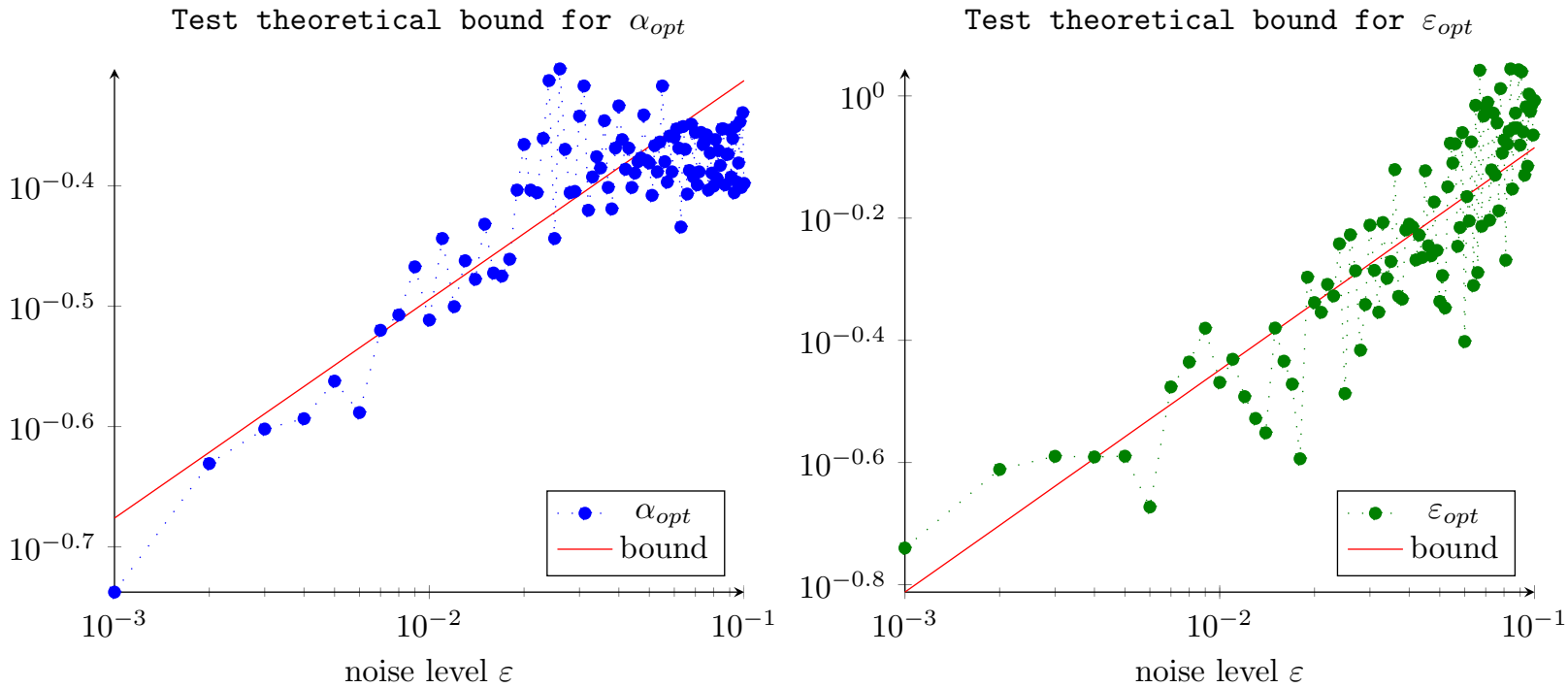


Figure : Comparison between the empirical results for α_{opt} (blue) and err_{opt} (green) and their relative bounds given by the Equations (1)-(2).