Penalized likelihood based estimation of systems of differential equations with applications

Javier González

University of Groningen
J. B. Institute of Mathematics and Computer Science
Systems Biology Centre for Metabolism and Ageing

ICL, BioMath seminar
November, 2012

Collaborators

Theoretical developements and algorithms (JBI)

Ernst Wit

Ivan Vujacic

Applications (SBC-EMA)

Barbara Bakker, Karen van Euen, Matthias Heinemman, Anne Mainena, Georges Janssens...

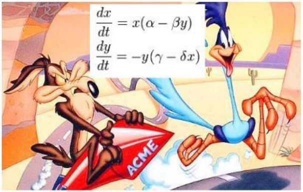
Dynamical systems and ODEs

Population dynamics

Excitable systems

Cell dynamics

Dynamical systems and ODEs



Population dynamics

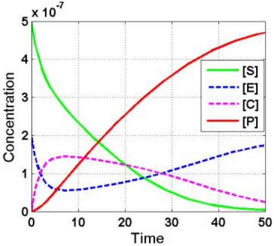
$$
\begin{aligned}
\frac{\partial V}{\partial t} &= \nabla^2 V + \frac{1}{\epsilon}(V - V^3/3 - W) \\
\frac{\partial W}{\partial t} &= \epsilon(V - \gamma W + \beta)
\end{aligned}
$$

Excitable systems

Cell dynamics

$$
\begin{aligned}
&k_{on}([E_T] - [ES])[S] = (k_{off} + k_{cat})[ES] \\
&k_{on}[E_T][S] - k_{on}[ES][S] = (k_{off} + k_{cat})[ES] \\
&k_{on}[E_T][S] = (k_{off} + k_{cat})[ES] + k_{on}[ES][S] \\
&[ES] = \frac{k_{on}[E_T][S]}{(k_{off} + k_{cat}) + k_{on}[S]} = \frac{[E_T][S]}{\left(\frac{k_{off} + k_{cat}}{k_{on}}\right) + [S]} \\
&\Rightarrow v = k_{cat}[ES] = \frac{k_{cat}[E_T][S]}{\left(\frac{k_{off} + k_{cat}}{k_{on}}\right) + [S]}
\end{aligned}
$$

Biology = Concentrations



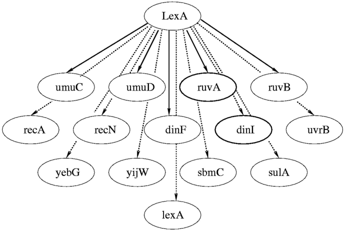
Modelling tool to understand biological systems:

Systems of differential equations

- Useful to model all type of decay processes, *e.g.* fluorescence, activated receptor returning to inactive state.
- Able to describe metabolic pathways.

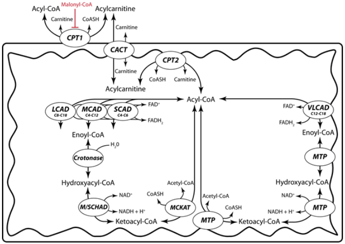
Biological Systems modelled by ODEs

E-Coli SOS system



- 14 genes regulated by a transcription factor.
- Particular kinetics parameters.

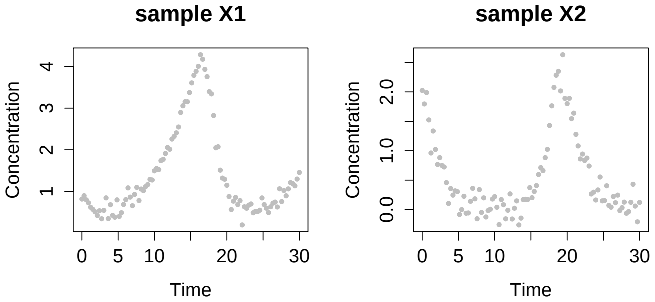
Fatty acid β-oxidation



- Kinetic equation for each enzyme.
- 45 ODEs and more than 200 parameters.

Problem statement: θ_1, θ_2, β_1 and β_2?

Data

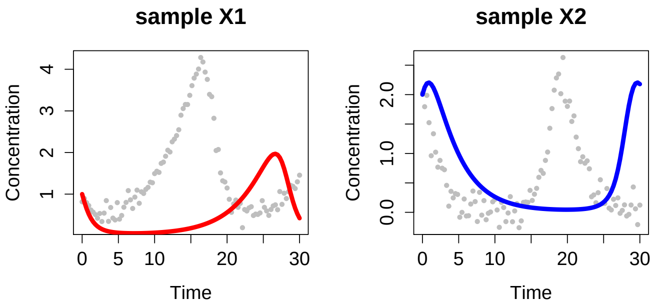


ODE model

$$\frac{d}{dt}x_1 = x_1(\theta_1 - \beta_1 x_2), \quad \frac{d}{dt}x_2 = -x_2(\theta_2 - \beta_2 x_1)$$

Problem statement: θ_1, θ_2, β_1 and β_2?

Data

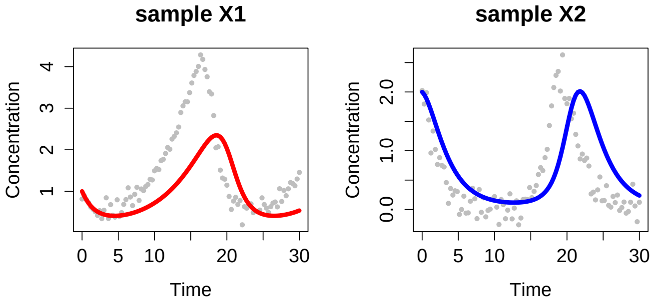


ODE model

$$\frac{d}{dt}x_1 = x_1(0.25 - 0.45x_2), \quad \frac{d}{dt}x_2 = -x_2(0.30 - 0.55x_1)$$

Problem statement: θ_1, θ_2, β_1 and β_2?

Data

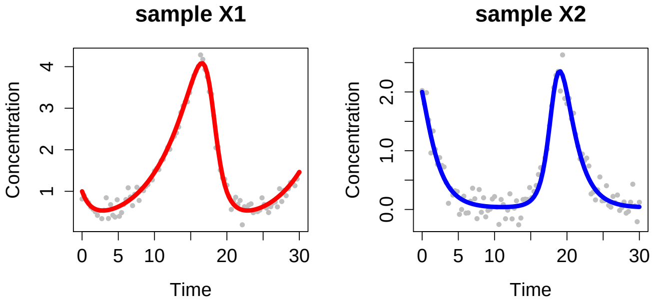


ODE model

$$\frac{d}{dt}x_1 = x_1(0.02 - 0.30x_2), \quad \frac{d}{dt}x_2 = -x_2(0.50 - 0.45x_1)$$

Solution: $\theta_1^* = 0.2$, $\theta_2^* = 0.35$, $\beta_1^* = 0.7$ and $\beta_2^* = 0.4$

Data



ODE model

$$\frac{d}{dt}x_1 = x_1(0.20 - 0.35x_2), \quad \frac{d}{dt}x_2 = -x_2(0.70 - 0.40x_1)$$

Issues in real applications

1. Large systems: Forward simulation requieres to solve the ODE which is computationaly expensive.
2. Lack of data: Small sample with a large uncertaintly.
3. True ODE model?
4. Unobserved components: Experimental limitations.

Example (E-Coli SOS system)

- 14 differential equations.
- 6 data points per gene.
- TF is unobserved.

Some previous approaches for ODE inference

Likelihood based approaches

- Estimation of the x_j's by nonparametric regression.
- Differentiation of $\hat{x}_j'$s and minimization over the parameters using a penalized likelihood *[Ramsay et al., 2007]*.

Other approaches

- Bayesian method. Solution of the ODE given as a Gaussian process *[Calderhead et al., 2008]*.
- Kernel Method for estimating 1-dimensional, periodic differential equations *[Steinke et al. 2008]*.

Our idea and approach

Idea

1. Combine the frequentist set-up with the kernel approach.

2. Parameter estimation: maximization of a likelihood with a Reproducing kernel Hilbert space (RKHS) based penalty.

3. EM algorithm to deal with missing data.

Steps

$$
\begin{array}{ccccccc}
\text{ODE} & \rightarrow & \text{RKHS} & \rightarrow & \text{Penalty} & \rightarrow & \text{Penalizaed likelihood} \\
dx = \theta x & \rightarrow & \mathcal{H} & \rightarrow & \Omega_H(x) & \rightarrow & \text{likelihood} + \Omega_H(x)
\end{array}
$$

Notation

ODE describing a dynamical system

$$P_{\boldsymbol{\theta}_j} x_j = f_{\beta_j}(x_1, \ldots, x_m, u_j), \;\; j = 1, \ldots, m.$$

Elements

- x_j, u_j: state variables and external forces defined on T.
- $P_{\boldsymbol{\theta}_j} = \sum_{k=0}^{d} \theta_{jk} D^k$ for $D^k = d^k/dt$ and $\theta_j = \{\theta_{j1}, \ldots, \theta_{jd}\}$.
- f_{β_j} known parametric function where $\beta_j = \{\beta_{j1}, \ldots, \beta_{jq}\}$.
- $S = \{(y_{ji}, t_i) \in \mathbb{R} \times T\}_{i,j=1}^{n,m}$: sample where $y_{ji} \sim \mathcal{N}(x_j(t_i), \sigma_j^2)$

Likelihood apprach

Constrained likelihood

$$l_j(\theta_j, \beta_j, \sigma_j | S_j) = -\frac{n}{2}\log(\sigma_j^2) - \frac{1}{2\sigma_j^2}\sum_{i=1}^{n}(y_{ji} - x(t_i))^2$$

for x_j satisfying that $P_{\theta_j} x_j = f_{\beta_j}$.

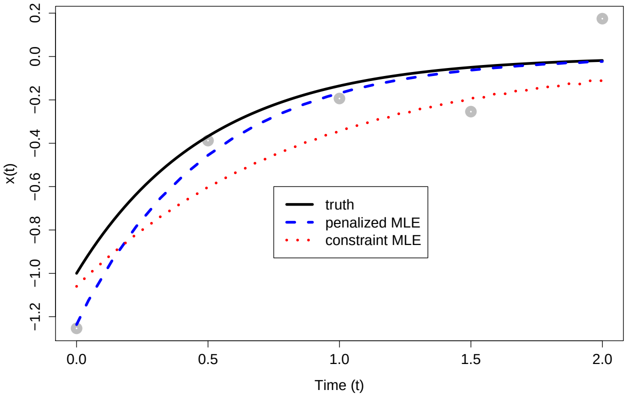
Penalized likelihood

$$l_{j,\lambda}(\theta_j, \beta_j, \sigma_j | S_j) = l_j(\theta_j, \beta_j, \sigma_j, x_j | S_j) + \lambda\Omega_j(x)$$

Questions: Why a penalized likelihood? How to define $\Omega_j(x)$?

MLE vs PMLE: $dx/dt = \theta x$ with $\theta_{true} = -2$

Explicit solution of the ODE is needed for MLE.



- This run: $\hat{\theta}_{MLE} = -1.12$, $\hat{\theta}_{PMLE} = -1.99$.
- 30 runs: $AD(\hat{\theta}_{MLE}) = 0.87$ and $AD(\hat{\theta}_{PMLE}) = 0.78$.

Defining $\Omega_j(x)$

- $P_{\boldsymbol{\theta}_j} x_j = 0$, generalization to non-homogeneous is feasible.
- P_{θ_j}, differential operator on some space of functions $\mathcal{H}$

$$\|x_j\|_{\mathcal{H}}^2 = \int_T (P_{\theta_j} x_j(t))^2 dt.$$

- When $\|x_j\|_{\mathcal{H}}^2 = 0$, x_j is a solution of $P_{\boldsymbol{\theta}_j} x_j = 0$.
- $\|x_j\|_{\mathcal{H}}^2 = 0$ is a convex functional.
- Use $\Omega_j(x) = \|x_j\|_{\mathcal{H}}^2$ as penalty.

Properties of $\Omega_j(x_j) = \|x_j\|^2_{\mathcal{H}}$

Reproducing kernel Hilbert spaces (RKHS) in a nutshell

- **Mercer kernel**: continous, symmetric and positive definite function $K : T \times T \to \mathbb{R}$.
- **RKHS**: completed space spanned by $x(t) = \sum_{i=1}^{n} \alpha_i K(t_i, t)$, where $n \in \mathbb{N}$, $t_i \in T$ and $\alpha_i \in \mathbb{R}$ and $\langle x_1, x_2 \rangle_{\mathcal{H}} = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \beta_j K(t_i, t_j)$.

Important properties

1. $\mathcal{H}$ is a RKHS whose reproducing kernel is a Green's function of $P_\theta^* P_\theta$: $P_\theta^* P_\theta K(t, z) = \delta(t - z)$.
2. $\|x\|^2_{\mathcal{H}} = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i^2 K(t_i, t_j)$.

Likelihood in the RKHS

Likelihood in the RKHS

$$l_{j,\lambda}(\theta_j, \beta_j, \sigma_j | S_j) = l_j(\theta_j, \beta_j, \sigma_j | S_j) + \lambda \boldsymbol{\alpha}_j^T \mathbf{K}_{\theta_j} \boldsymbol{\alpha}_j$$

- $(\mathbf{K}_{\theta_j})_{is} = K_{\theta_j}(t_i, t_s)$.
- $\boldsymbol{\alpha}_j \in \mathbb{R}^n$ characterizes x_j.
- Note: $(x_j(t_1), \ldots, x_j(t_n))^T = \mathbf{K}_{\theta_j} \boldsymbol{\alpha}_j$

Parameters estimates

$$(\hat{\Theta}, \hat{B}, \hat{\Sigma} | S) = \arg \max_{\Theta, B, \Sigma} \sum_{j=1}^{m} l_{j,\lambda}(\theta_j, \beta_j, \sigma_j | S_j)$$

Problem: function K_θ is rarely available

- Replace $\boldsymbol{\alpha}_j^T \mathbf{K}_{\theta_j} \boldsymbol{\alpha}_j$ by an approximation $\boldsymbol{\alpha}_j^T \tilde{\mathbf{K}}_{\theta_j} \boldsymbol{\alpha}_j$.
- $\mathbf{P}_{\theta_j} = \sum_{k=0}^{d} \theta_{jk} \mathbf{D}^k$: difference operator defined on $t_1, \ldots, t_n$

$$\mathbf{D} = \Delta^{-1} \cdot \begin{pmatrix} -1 & 1 & & & \\ -1 & 0 & 1 & & \\ & & & \ddots & \\ & & & -1 & 1 \end{pmatrix}$$

where $\Delta = diag(t_2 - t_1, t_4 - t_2, \ldots, t_n - t_{n-2}, t_n - t_{n-1})$.

- Focus on the difference equation and use the approximation of $\mathbf{K}_{\theta_j}$ given by

$$\tilde{\mathbf{K}}_{\theta_j} = (\mathbf{P}_{\theta_j}^T \mathbf{P}_{\theta_j})^{-1}.$$

Error of the finite dimensional approximation

$$\frac{d}{dt}x_j(t_{i-1}) = \lim_{t_i \to t_{i-1}} \frac{x_j(t_i) - x_j(t_{i-1})}{t_i - t_{i-1}} \approx \frac{x_j(t_i) - x_j(t_{i-1})}{t_i - t_{i-1}}$$

Idea

- To include a number of hidden data points $(\mathbf{t}_H^*, \mathbf{y}_H^*)$.
- K_{θ_j} only depends on the $t_i's$ (so, we can do it!).
- More points $\rightarrow$ better approximations of the derivatives.

Expectation-Maximization algorithm

- S_0, observed data.
- S_H, data set of hidden points $(\mathbf{t}_H^*, \mathbf{y}_H^*)$.
- Parameters, $\Gamma = \{\Theta, B, \Sigma\}$. Start with some $\Gamma^{(0)}$.

E-step

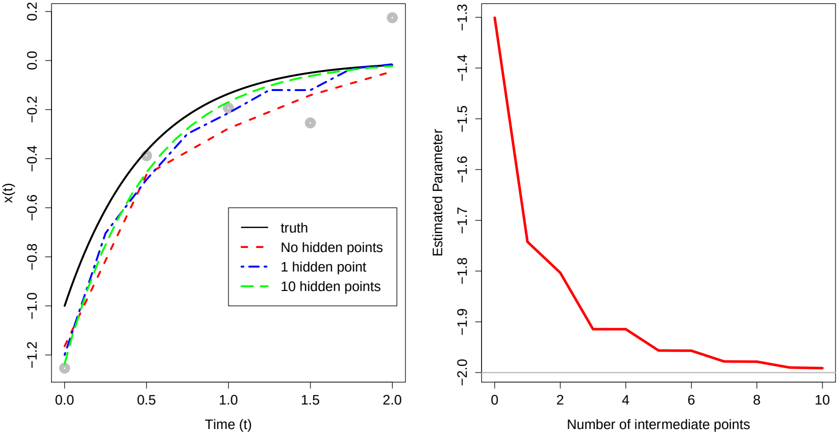
$$Q(\Gamma|\Gamma^{(t)}) = E_{S_H|S_O,\Gamma^{(t)}}\left[l_\lambda(\Gamma|S_O, S_H)\right]$$

where l_λ is the likelihood for the whole system.

M-step

$$\Gamma^{(t+1)} = arg\max_{\Gamma} Q(\Gamma|\Gamma^{(t)})$$

Performance of the EM approach



Non-homogeneous cases $P_\theta x = f_\beta$

Penalized likelihood

$$l_{\lambda,j}(\theta_j, \beta_j, \sigma_j | S_j) = -\frac{1}{2\sigma_j^2} \sum_{i=1}^{n} (y_{ji} - x(t_i))^2 - \lambda \| P_{\theta_j} x_j - f_\beta \|^2$$

Transformation

$$\tilde{x}_j = x_j - P^{-1} f_\beta$$

$$\tilde{y}_{ji} = y_{ji} - P^{-1} f_\beta(t_i)$$

New Penalized likelihood

$$l_{\lambda,j}(\theta_j, \beta_j, \sigma_j | S_j) = -\frac{1}{2\sigma_j^2} \sum_{i=1}^{n} (\tilde{y}_{ji} - \tilde{x}(t_i))^2 - \lambda \| P_{\theta_j} \tilde{x}_j \|^2$$

We have seen so far...

- To regularize the likelihood is a good idea. Biased but more robust estimates.
- The constrained likehood can be written like a penalized likelihood in a RKHS.
- The Green's function of the differential operator is the key ingredient.
- The finite dimensional representation of the problem allows to use this idea in general cases.
- A EM formulation allows to correct errros derived from the finite dimensional aproximation of the ODE.

Implementation

Package ‘oderkhs’

November 1, 2012

Type Package

Title Reproducing kernel Hilbert Space based estimation of parameters of systems of Ordinary Differential equations

Version 1.0

Date 2012-10-31

Author Ivan Vujacic `<i.vujacic@rug.nl>`, Javier Gonzalez `<j.gonzalez.hernandez>`, `Ernst Wit <e.c.witg@rug.nl>`

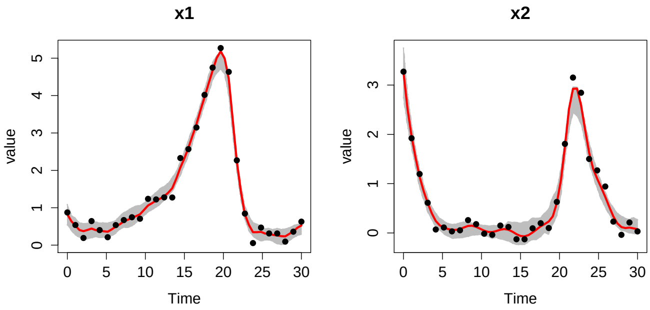
Depends R (>= 2.10.1), expm, pspline, magic, MASS, corpcor, pspline, gplots

Maintainer Ivan Vujacic `<i.vujacic@rug.nl>`

Description These functions implement RKHS based estimation of parameters of ODEs. They provide parameter estimates, confidence intervals and estimates of state variables.

'oderkhs' in our original example (0.2,0.35,0.7,0.4)

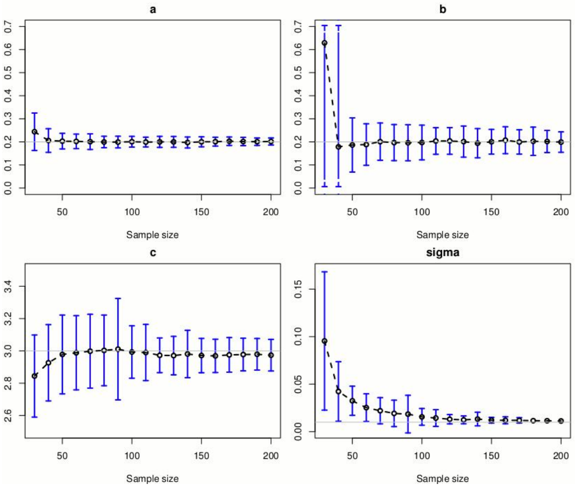
```
> res = rkhs(data,f,coef,times,lambda,int=1,em.it=5)
> res.boot = ci.boot(res.ode,50)
> plot.odesol(res.boot)
```




```
> res.boot$ci
2.5% 0.15 0.32 0.47 0.24
97.5% 0.29 0.41 0.73 0.45
```

FitzHugh-Nagumo model ($a, b = 0.2$ and $c = 3$)

$$\frac{dx_1}{dt} = c\left(x_1 - \frac{x_1^3}{3} + x_2\right), \frac{dx_2}{dt} = \frac{1}{c}(x_1 - a + bx_2)$$



Comparative with the method proposed in (Ramsay et al., 2009)

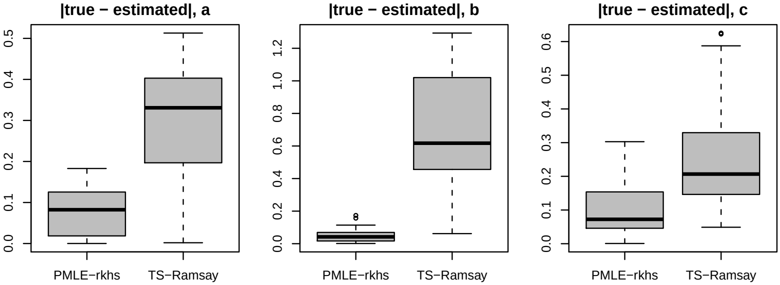
- We compare the estimates for sample sizes $n = 50$ and $n = 300$.
- 50 replicates of the experiment.
- Differences between the estimates of the parameters and their true values are used across the 50 replicates are considered.
  - Maximum error.
  - Averaged error.
  - Minimum error.

Errors comparative with (Ramsay et al., 2009)

n	FHN pars.	Av. error PMLE	Av. error Rams.	Max. error PMLE	Max. error Rams.	Min. error PMLE	Min. error Ram.
50	a	*0.0780*	0.2892	*0.1829*	0.5130	*0.0000*	0.0019
	b	*0.0485*	0.7023	*0.1737*	1.2936	*0.0012*	0.0619
	c	*0.0983*	0.2503	*0.3028*	0.6253	*0.0006*	0.0488
300	a	0.0058	*0.0032*	0.0150	*0.0126*	0.0000	0.0000
	b	0.0175	*0.0101*	0.0609	*0.0318*	0.0005	*0.0002*
	c	0.0348	*0.0134*	0.1133	*0.0476*	0.0008	*0.0010*

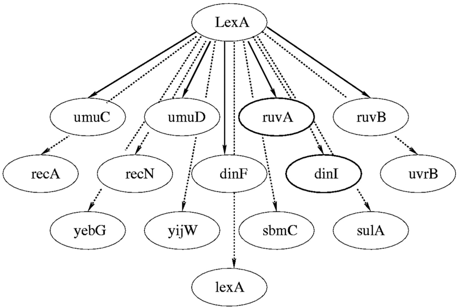
- For $n = 50$ the proposed methodology is the best under the three criteria.
- For $n = 300$ only minor differences are found.

Errors comparative with (Ramsay et al., 2009)



The proposed method outperform (Ramsay et al., 2009) in small-sample scenarios.

E-Coli SOS system



- Model

$$\dot{x}_k(t) = \beta_k \frac{1}{\gamma_k + \eta(t)} + \varphi_k - \delta_k x_k(t),$$

- Gene-dependent kinetics parameters: $\beta_k, \gamma_k, \delta_k$ and φ_k.

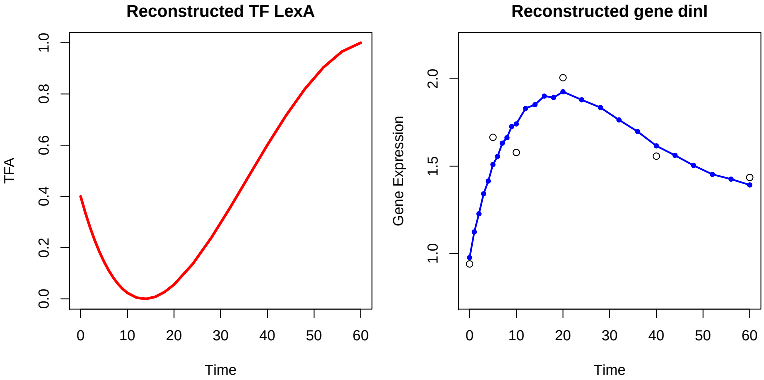
E-Coli SOS system: Data

- Data set of 14 expression genes.
- The genes are targets of the master repressor LexA and their expression is studied under UV exposure (40J/m2).
- Abundance of the mRNA molecules measured at 0, 5, 10, 20, 40 and 60 minutes.
- The LexA abundance is unobserved. Is modelled by a splines expansion

$$\eta(t) = \sum_{k=1}^{d} \mu_k \phi_k.$$

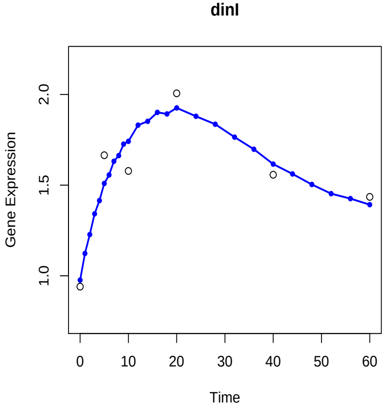
E-Coli SOS system: Results

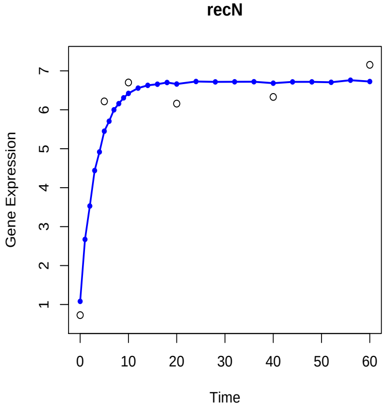
Reconstructed the activity of the repressor $\eta(t)$ gene dinI.



Confidence intervals for the parameters $CI_{95\%}(\beta) = (0.56, 1.39)$ $CI_{95\%}(\delta) = (0.15, 0.27)$ $CI_{95\%}(\gamma) = (1.71, 2.83)$, $\sigma = 0.04$.

E-Coli SOS system: Genes behaviour





- The activity of gene dinI decline after minute 20.
- The activity of gene recN does not decline.

E-Coli SOS system Parameters estimates and CI

Gene	$\hat{\beta}_k$	$CI_{95\%}(\beta_k)$	$\hat{\delta}_k$	$CI_{95\%}(\delta_k)$	$\hat{\gamma}_k$	$CI_{95\%}(\gamma_k)$
recN	6912.05	(6465.5, 6933.1)	0.29	(0.21, 0.30)	3568.3	(3527.4, 4433.5)
umuC	79.94	(78.22, 88.71)	0.09	(0.09, 0.10)	123.21	(117.45, 124.12)
ijW	12.44	(2.90, 28.20)	0.42	(0.21, 0.46)	22.54	(9.05, 76.17)
ruvA	5.27	(3.82, 6.30)	0.42	(0.35, 0.45)	6.02	(4.64, 7.15)
lexA	4.75	(2.28, 7.56)	0.33	(0.25, 0.38)	7.70	(4.23, 11.75)
sulA	4.36	(1.03, 9.24)	0.39	(0.15, 0.40)	2.66	(1.42, 6.88)
umuD	2.99	(2.13, 3.94)	0.15	(0.13, 0.16)	5.11	(4.08, 6.66)
yegG	2.10	(1.79, 2.42)	0.23	(0.21, 0.25)	3.72	(3.31, 4.20)
ruvB	1.98	(1.79, 2.23)	0.27	(0.26, 0.29)	4.38	(4.11, 4.74)
uvrB	1.89	(1.26, 3.40)	0.10	(0.09, 0.11)	7.11	(5.07, 12.08)
dinI	0.99	(0.56, 1.39)	0.21	(0.15, 0.27)	2.36	(1.71, 2.83)
recA	0.48	(0.39, 0.67)	0.11	(0.10, 0.13)	0.85	(0.72, 1.11)
sbmC	0.26	(0.22, 0.30)	0.11	(0.11, 0.12)	0.71	(0.62, 0.78)
dinF	5.01	(2.06, 32.16)	0.12	(0.08, 0.18)	31.59	(17.32, 167.07)

Genes *recN* and *umuC* do not decline after minute 20. Miss specified model (no regulation).

$$\dot{x}_k(t) = \varphi_k + \delta_k x_k(t),$$

Conclusions

- Penalized likelihood approach where the ODE is directly used to define the penalty. The Green function of the differential operator is the key ingredient.
- Efficient approach to estimate the parameters of ODE.
- Able to deal with the lack of data common in most real applications. Specially accurate in small sample scenarios.
- Available online within the CRAN project.

Work in progress / future work

Methodological:

- Explore statistical properties of the method.
- More efficient estimation of the parameters.
- Generalization to stochastic ODEs.

Applied:

- Statistical analysis of competition among enzymes in Fatty acid β oxidation. 45 ODEs with more than 200 parameters. Impact on ageing in mices.

- Analysis of the “ageing system" in yeast. Combine this method with other methodologies to get a high level understanding of the ageing process in yeast.

Thanks!

Acknowledgements:
Project SBC-EMA-435065