# Learning parameters in ODEs

## Application to biological networks

Florence d'Alché-Buc

**Joint work with Minh Quach and Nicolas Brunel**
IBISC FRE 3190 CNRS, Université d'Évry-Val d'Essonne, France

# Nonlinear State-Space Model based on ODEs

## Continuous time ODE model

$$\begin{aligned} \frac{d\mathbf{x(t)}}{dt} &= \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t); \theta) \\ \mathbf{y}(t) &= \mathbf{H}(\mathbf{x}(t), \mathbf{u}(t); \theta) + \epsilon(t) \end{aligned}$$

## The augmented system at discrete-time points $t_0, t_1, ..., t_{N-1}$ possibly irregularly spaced

- $$\begin{aligned} \theta(t_{k+1}) &= \theta(t_k) \\ \mathbf{x}(t_{k+1}) &= \mathbf{F}(\mathbf{x}(t_k), \mathbf{u}(t_k); \boldsymbol{\theta_k}) \\ \mathbf{y}(t_k) &= \mathbf{H}(\mathbf{x}(t_k), \mathbf{u}(t_k); \boldsymbol{\theta_k}) + \epsilon(t_k) \end{aligned}$$
- with
$$\mathbf{F}(\mathbf{x}(t_k), \mathbf{u}; \boldsymbol{\theta_k}) = \mathbf{x}(t_k) + \int_{t_k}^{t_{k+1}} \mathbf{f}(\mathbf{x}(\tau), \mathbf{u}(\tau); \boldsymbol{\theta_k}) d\tau$$
- $u(.)$ is a control variable

# Comments on this model

- The hidden process is no more noisy (for sake of simplicity)
- Timepoints can be irregularly spaced due to the integration (numerical)
- Even if $f(.;\theta)$ is linear, the hidden process is not linear

# Parameter and hidden variable estimation (1)

## Parameter estimation

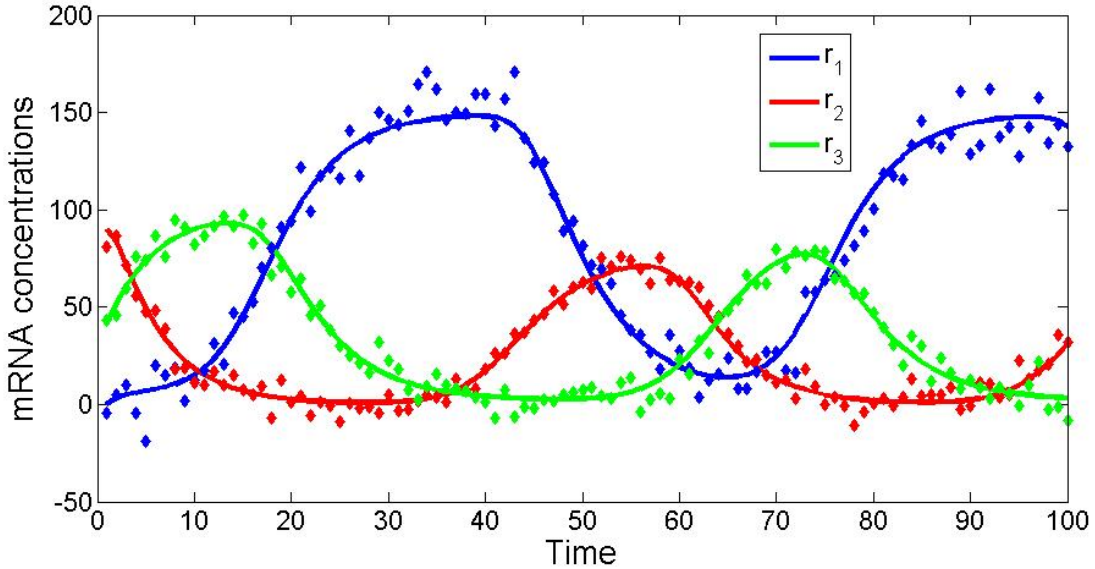
- According the nature of function $f$, the graph of interactions between components may be encapsulated in $f$ or not
- In the next slides, we assume that the graph structure is given
- Current work includes coupling parameter estimation algorithms with stochastic search for the graph of interactions

## Hidden state estimation

- For instance in Michaelis-Menten equations and Hill kinetics, recover proteins concentrations from gene expression observation

# Parameter and hidden variable estimation (2)

- Prior on $x(t_0)$ and prior on $\theta$
- Learning as filtering (or smoothing) with an augmented approach
- Random initializations


mRNA concentrations
Time
r1
r2
r3
200
150
100
50
0
-50
0
10
20
30
40
50
60
70
80
90
100

# Estimation of parameters

Table: $T = 25$ observations. Average with std for 100 samples, each used for 50 random initialisations: case of Repressilator

| Parameter | True Parameter | UKF estimation | UKS estimation |
|---|---|---|---|
| $v_1$ | 150 | $150.50 \pm 4.14$ | $151.49 \pm 4.51$ |
| $v_2$ | 80 | $76.80 \pm 4.91$ | $81.45 \pm 4.09$ |
| $v_3$ | 100 | $118.01 \pm 17.2$ | $98.41 \pm 7.45$ |
| $k_{1,2}$ | 50 | $51.27 \pm 3.63$ | $46.35 \pm 3.44$ |
| $k_{2,3}$ | 30 | $31.21 \pm 2.81$ | $29.07 \pm 2.64$ |
| $k_{3,1}$ | 40 | $37.8 \pm 2.95$ | $39.71 \pm 2.85$ |

# Parameter Estimation

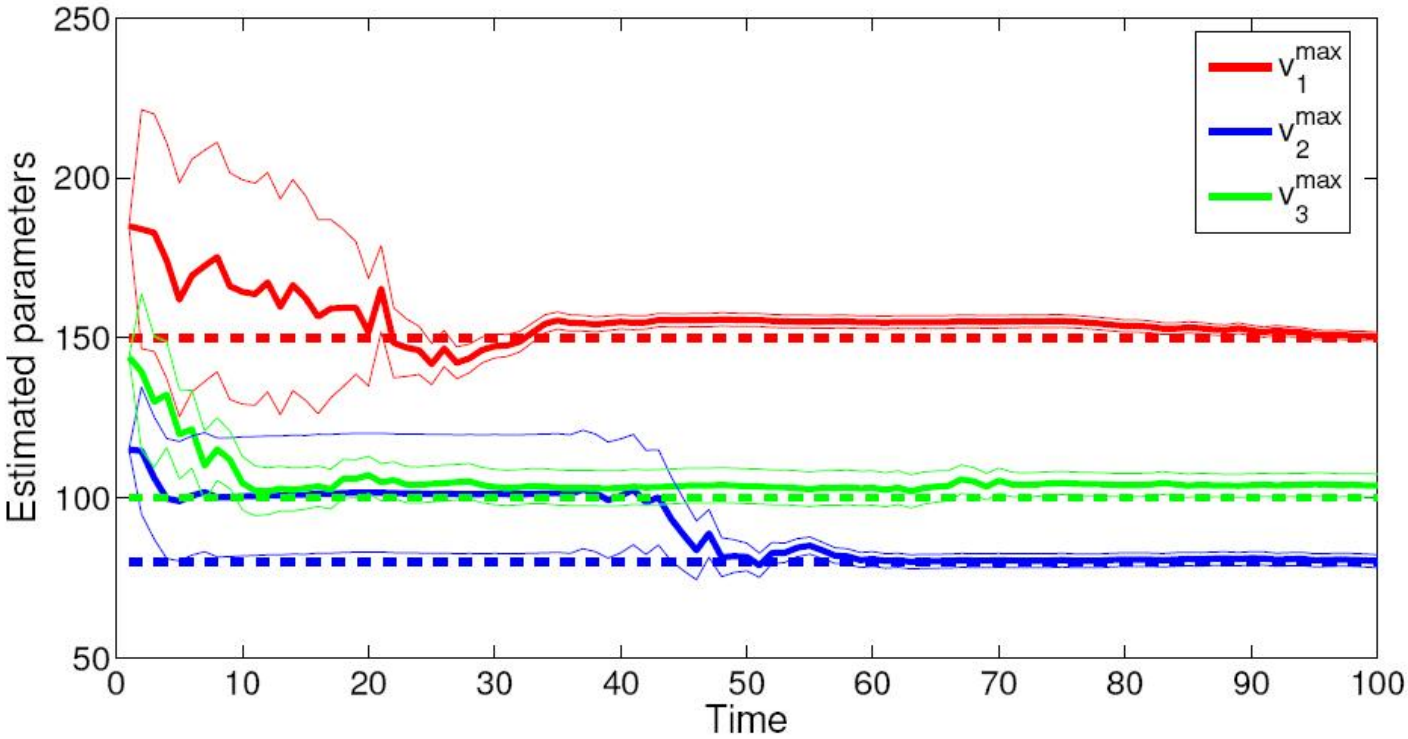



**Fig. 4.** Recursive estimation of the maximal rate of Michaelis-Menten kinetics through time for the case $S = 1$ and sampling time $\Delta_t = 0.2$ (corresponds to 100 data points). Dash lines: true parameters. Solid lines: Estimated parameters along with their confidence intervals

# Parameter Estimation

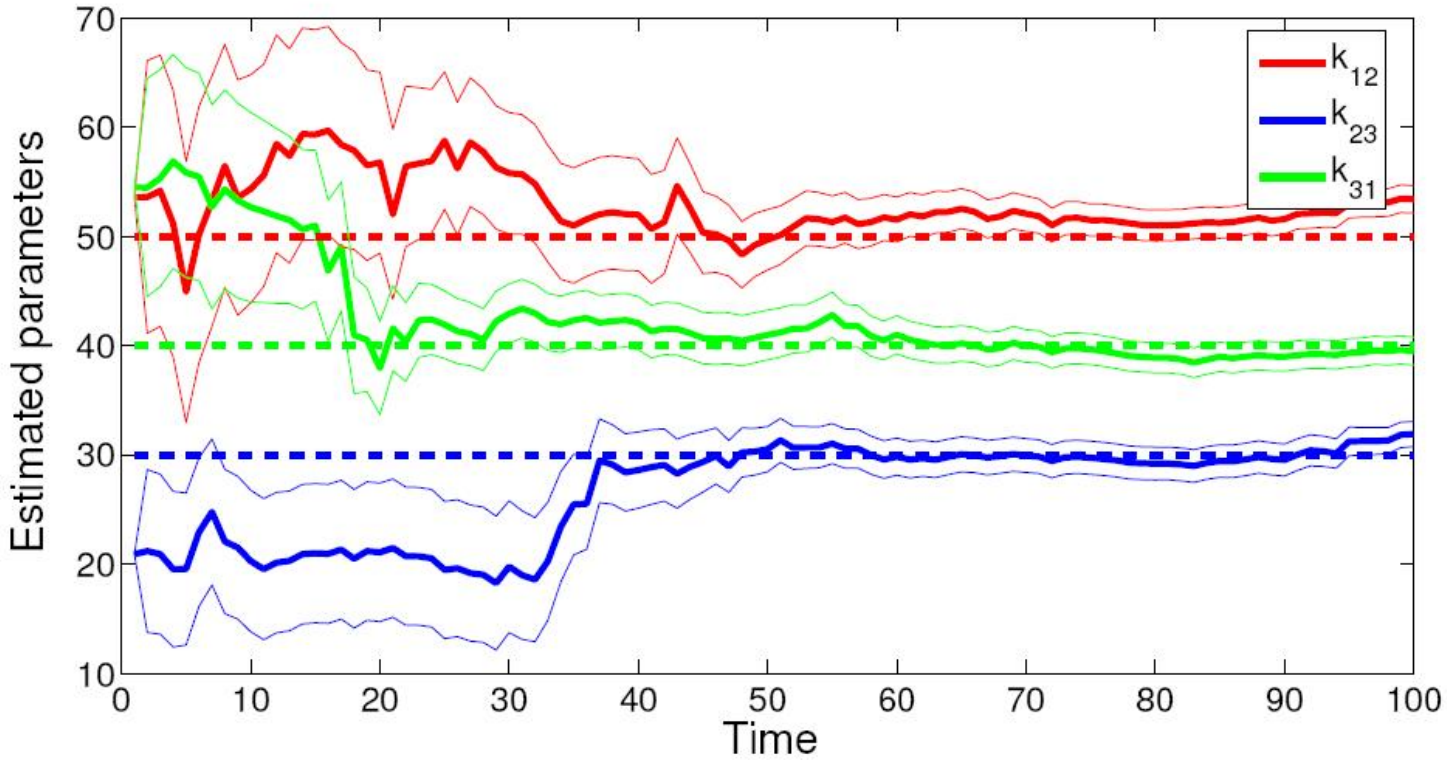



**Fig. 5.** Recursive estimation of Michaelis constants $k_{12}, k_{23}, k_{31}$ through time.

# State Estimation

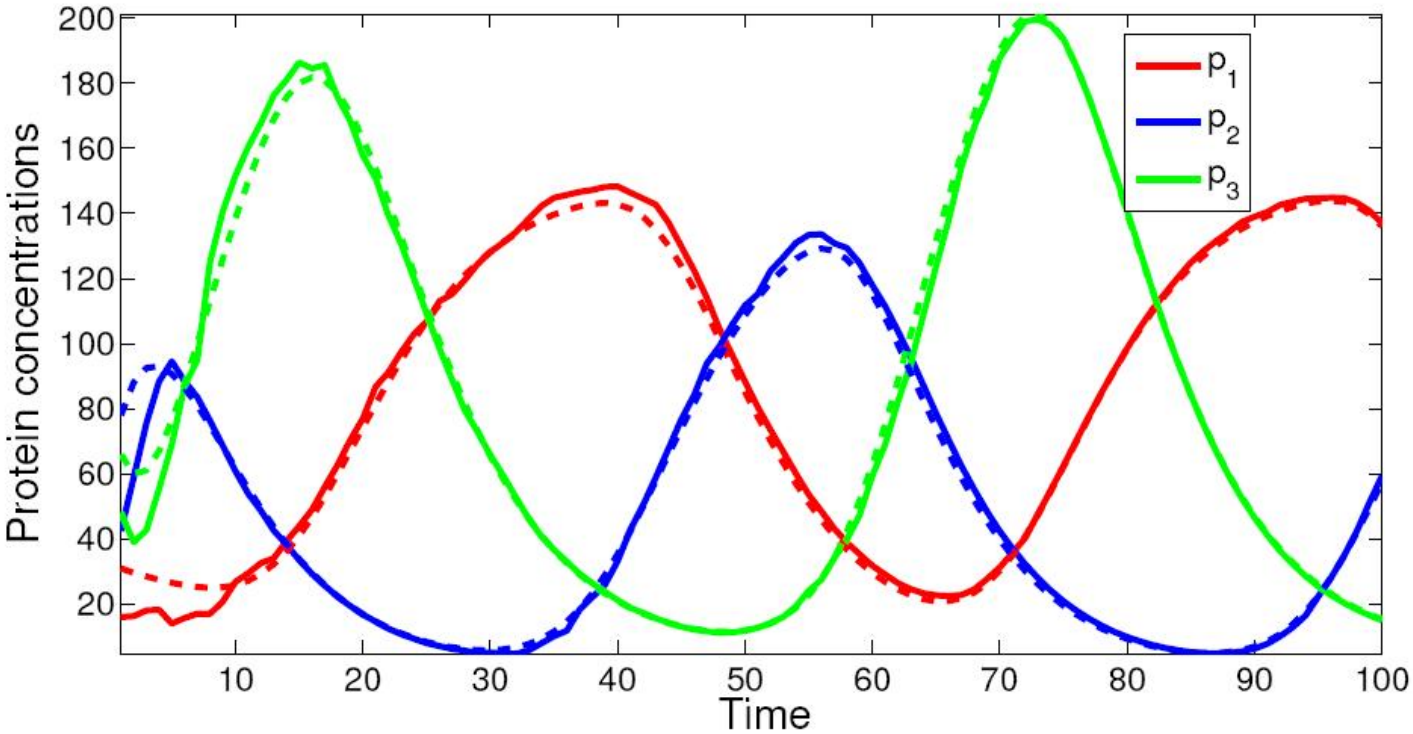
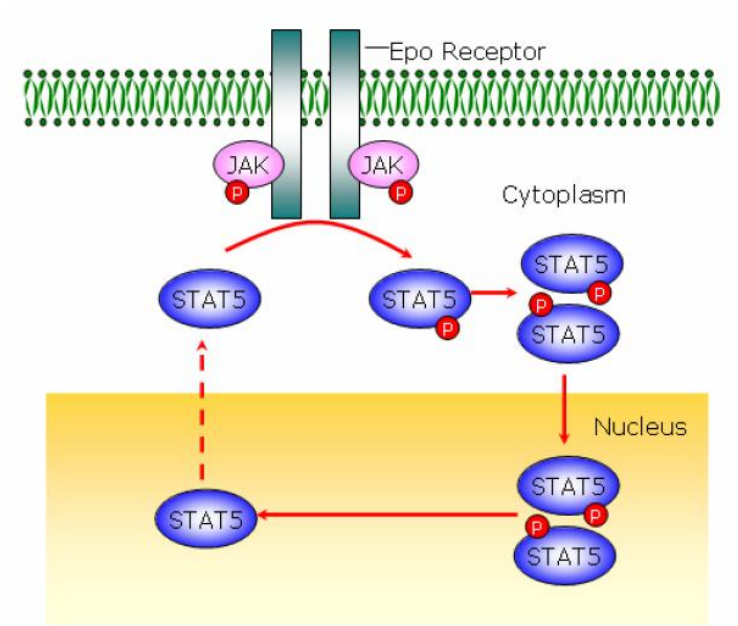



**Fig. 3.** The evolution of the true (dashed) and estimated (solid) protein concentrations.

# JAK-STAT signaling pathway




- ODE:

$$\begin{cases} \dot{x}_1(t) &= -a_1 x_1(t) u(t) + 2a_4 x_4(t) 1_{\{t \geq \tau\}} \\ \dot{x}_2(t) &= a_1 x_1(t) u(t) - 2a_4 x_2^2(t) \\ \dot{x}_3(t) &= -a_3 x_3(t) + x_2^2(t) \\ \dot{x}_4(t) &= a_3 x_3(t) - a_4 x_4(t) 1_{\{t \geq \tau\}} \end{cases}$$

- Observed variables

$$\begin{aligned} y_1 &= x_2 + 2x_3 \\ y_2 &= (x_1 + x_2 + 2x_3) \end{aligned}$$

- Experimental data: 16 time points
- $\theta = (a_1, a_3, a_4)^\top$ is the parameters to be estimated

# Convergence of the prediction error

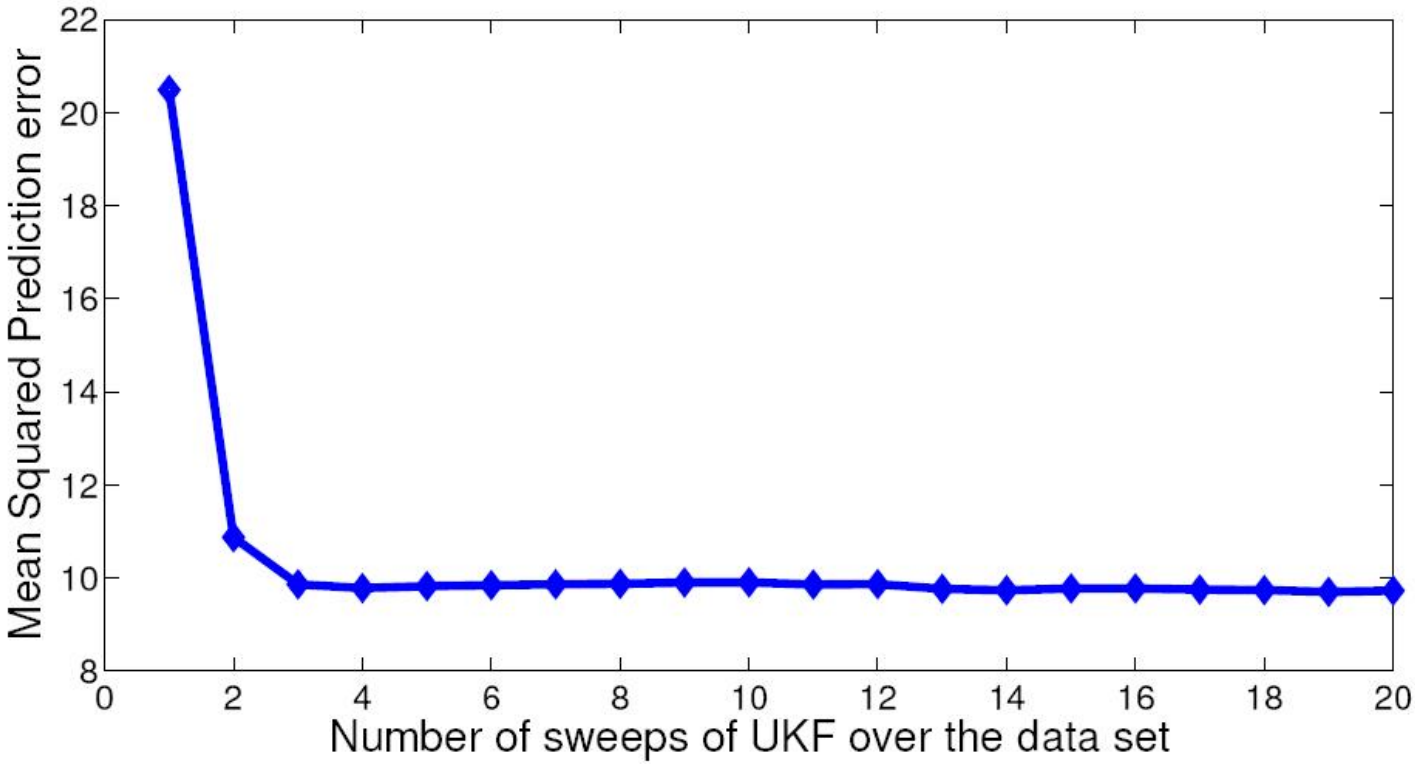



**Fig. 11.** Convergence of the prediction error when applying multiple UKF sweeps over the data set

# Prediction vs Experimental data

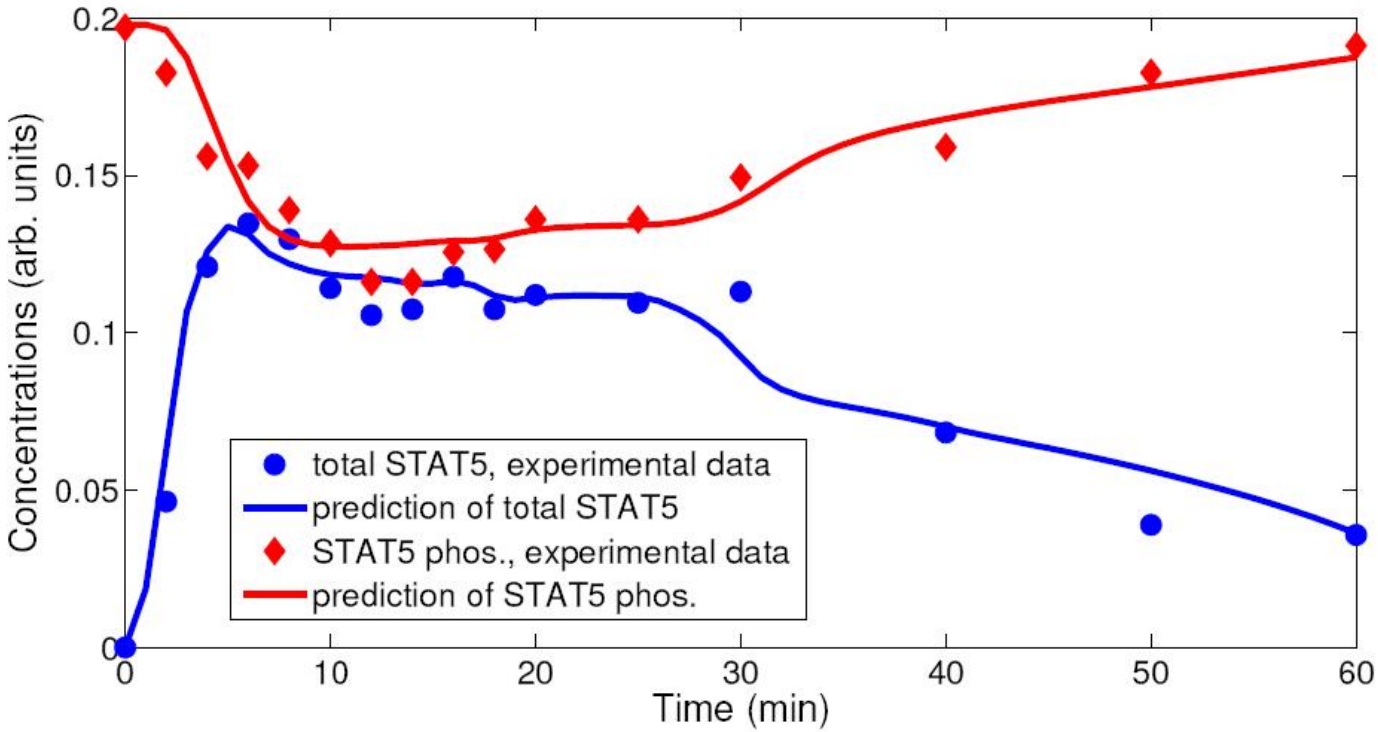



**Fig. 12.** Prediction of STAT5 phosphorylation and total amount of STAT5.

# Prediction of state variables

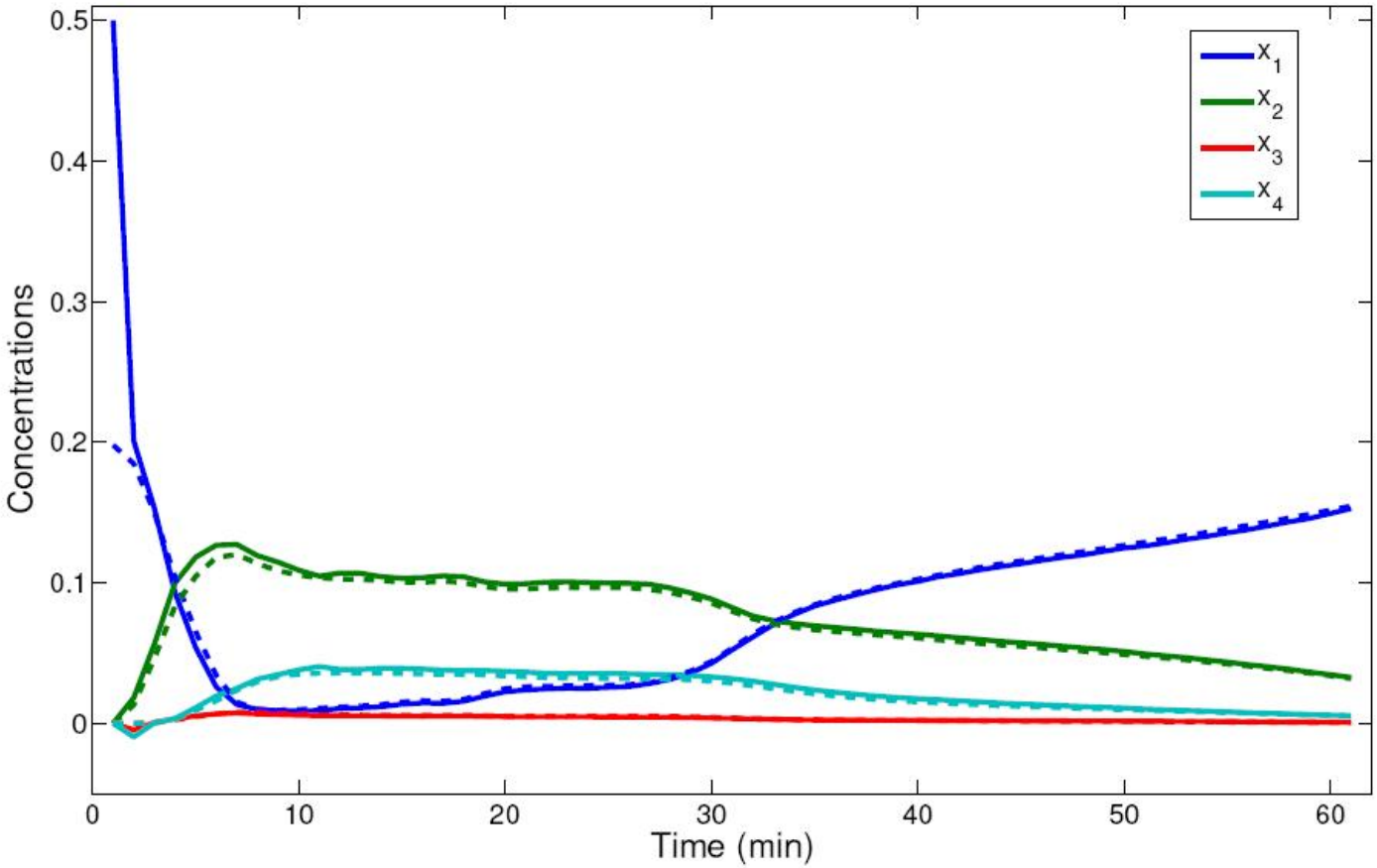



**Fig. 13.** The evolution of the true (dashed) and estimated (solid) concentrations of the four unobserved variables.

# Challenges in (dynamical) modeling approaches

- Identifiability of dynamical models$\longrightarrow$ simpler models : nonparametric models
- Prior knowledge
- Scaling to large networks by mixture models
- Non stationarity by switching models (see Xing's team work in ISMB 2009 and ICML 2009, linear models)
- Coupling parameter and structure learning
- **coupled systems** : metabolic and regulatory networks, protein-protein interactions and regulatory network