Réunion satellite VicAnne
Conférence JOBIM
Marseille 2007

Identification Procedure for PWA Models of Genetic Regulatory Networks

Samuel Drulhe
Samuel.Drulhe@inrialpes.fr
R. Porreca, H. de Jong, G. Ferrari-Trecate

INRIA Rhône-Alpes
655, avenue de l'Europe
38330 Montbonnot Saint-Martin, France

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

UNIVERSITE JOSEPH FOURIER
SCIENCES.TECHNOLOGIE.SANTÉ

Overview

- Genetic regulatory networks in brief
- Identification for PWA models of genetic regulatory networks
- PWA system identification
- Description of our approach
- Reconstruction of switching thresholds
- A case study: carbon starvation response of *E. coli*
- Conclusions

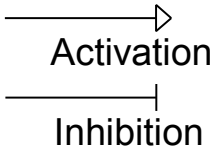
Overview

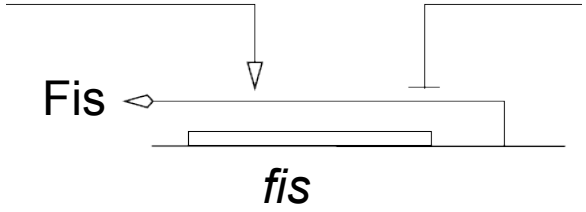
- Genetic regulatory networks in brief
- Identification for PWA models of genetic regulatory networks
- PWA system identification
- Description of our approach
- Reconstruction of switching thresholds
- A case study: carbon starvation response of *E. coli*
- Conclusions

Genetic regulatory networks

- **Genetic regulatory networks** underlie functioning and development of living organisms
  Components: genes, proteins, metabolites, and their mutual regulatory interactions

Genes



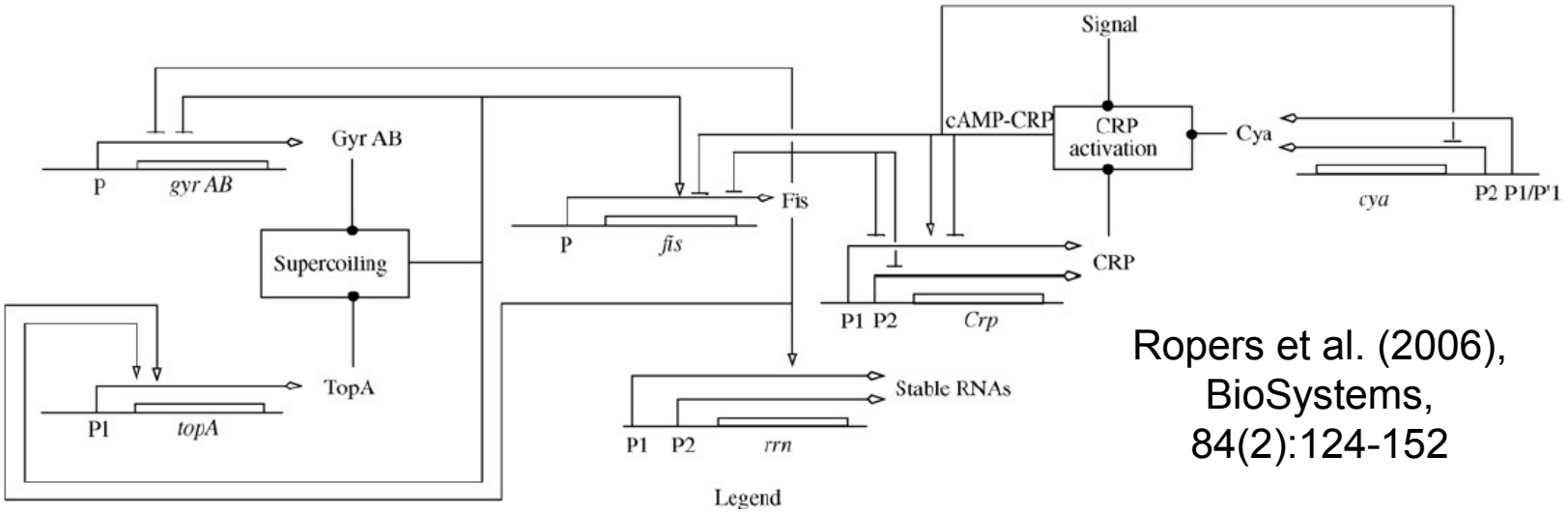


- Gene: dynamical system coding for a molecule (e.g. a protein)
- Genes are regulated by the concentration of proteins present in the cell
  - Genes can be turned on and off

Genetic regulatory networks

- GRN are usually **large** (many genes) and **complex** (feedback loops)

GRN governing *E. coli* carbon starvation response

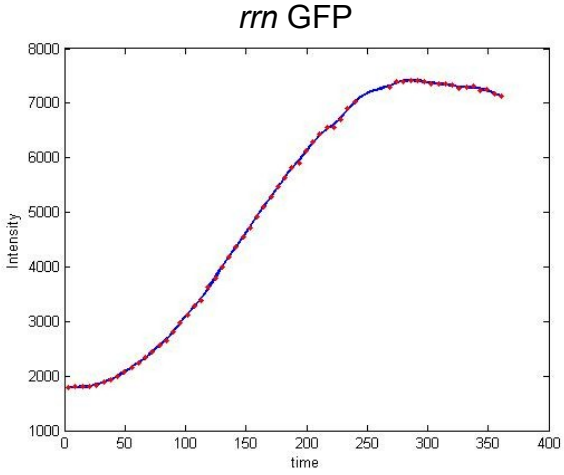


Ropers et al. (2006), BioSystems, 84(2):124-152

Expression data

Experimental techniques in biology have led to the production of enormous amount of data on the dynamics of gene expression:

- DNA microarrays
- Gene reporter systems



Time-series measurement of fluorescence or luminescence

Overview

- Genetic regulatory networks in brief
- Identification for PWA models of genetic regulatory networks
- PWA system identification
- Description of our approach
- Reconstruction of switching thresholds
- A case study: carbon starvation response of *E. coli*
- Conclusions

Data-driven modeling of GRNs

System identification problem: derive a model of the regulatory interactions according to measurements and model structure

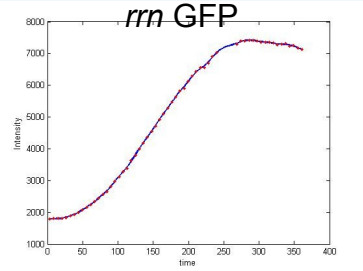
List of:

- genes
- proteins
- small molecules

List of:

- genetic interactions
- dynamical parameters

Expression data



Model structure

$$\frac{dx}{dt} = f(x) - \gamma(x)x$$

State of the art

Classes of dynamical models that were used for biological network identification:

x : a vector of protein concentrations

f, γ : functions that map interactions

$$\frac{dx}{dt} = f(x) - \gamma(x)x$$

synthesis rate ≥ 0 degradation rate > 0

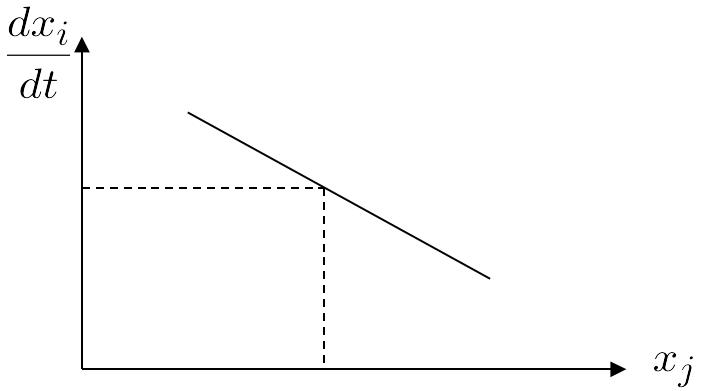
State of the art

Classes of dynamical models that were used for biological network identification:

- Linear (Gardner et al., Science 301 (2003) 102–105)
  → only valid near an equilibrium point

$$\frac{dx_i}{dt} = f_i(x) - \gamma_i x_i$$

affine

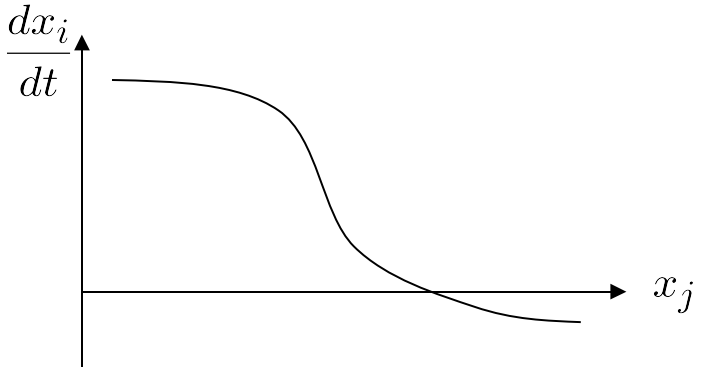


State of the art

Classes of dynamical models that were used for biological network identification:

- Linear (Gardner et al., Science 301 (2003) 102–105)
  → only valid near an equilibrium point
- Nonlinear (Jaeger et al., Nature 430 (2004) 368–371)
  → more adequate description but difficult to use for identification

$$\frac{dx_i}{dt} = f_i(x) - \gamma_i(x)x_i$$



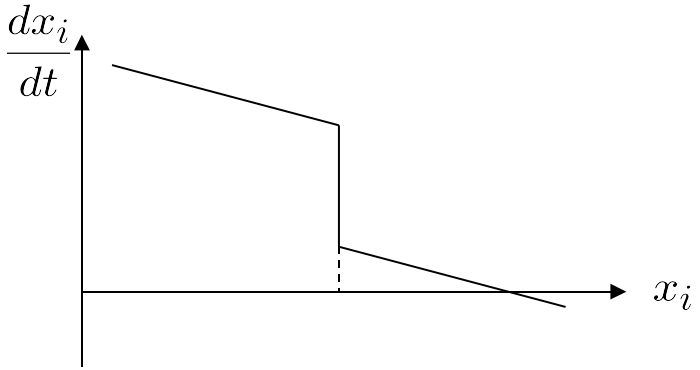
State of the art

Classes of dynamical models that were used for biological network identification:

- Linear (Gardner et al., Science 301 (2003) 102–105)
  → only valid near an equilibrium point
- Nonlinear (Jaeger et al., Nature 430 (2004) 368–371)
  → more adequate description but difficult to use for identification
- Piecewise Affine
  → compromise between linear and non-linear

$$\frac{dx_i}{dt} = f_i(x) - \gamma_i(x)x_i$$

piecewise constant



State of the art

Introduced by Glass and Kauffman in the 1970s

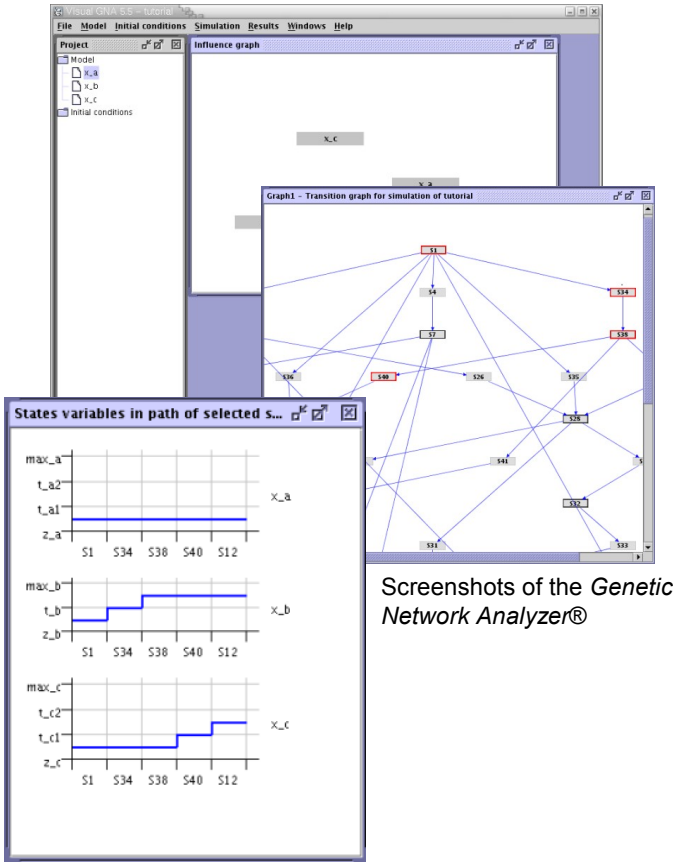
Piecewise Affine

→ compromise between linear and non-linear

- deJong et al., Bull. Math. Biol. 66 (2004) 301–340
- Belta et al., HSCC04, Vol. 2993 of LNCS (2004) 111-125
- Ghosh and Tomlin, Syst. Biol. 1 (2004) 170–183
- Batt et al., HSCC05, Vol. 3414 of LNCS (2005) 134–150

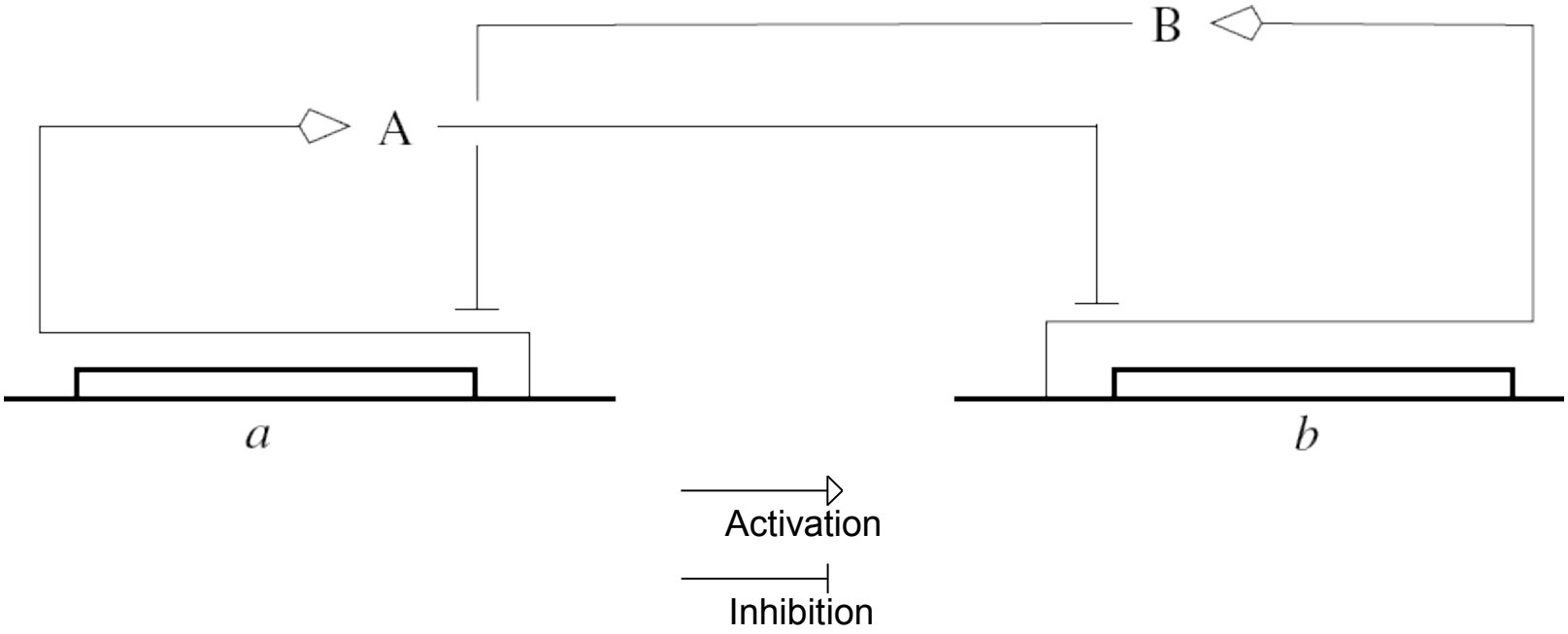
→ abstraction techniques available

→ identification methods for PWA systems available

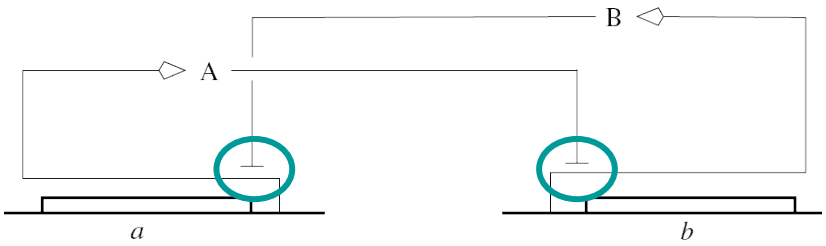


Screenshots of the *Genetic Network Analyzer®*

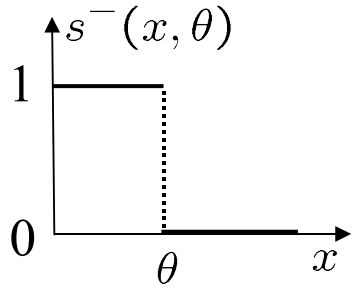
PWA models: a simple example.



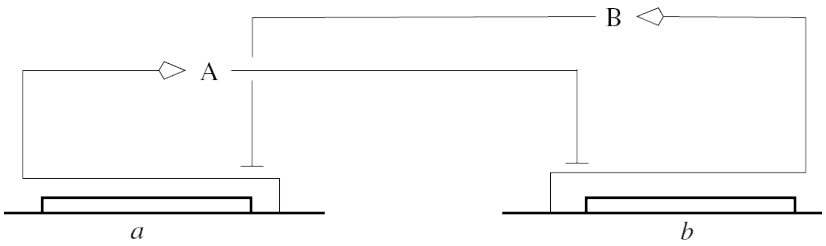
PWA models: a simple example.



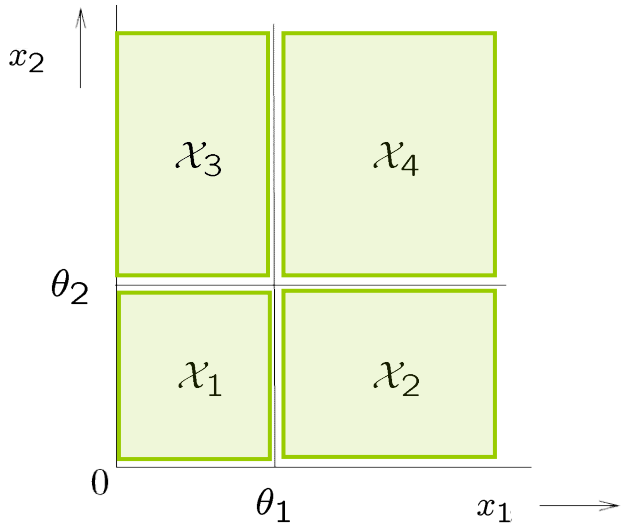
$$\dot{x}_1 = \kappa_1 s^-(x_2, \theta_2) - \gamma_1 x_1$$
$$\dot{x}_2 = \kappa_2 s^-(x_1, \theta_1) - \gamma_2 x_2$$



PWA models: a simple example.



$$\dot{x}_1 = \kappa_1 \, s^-(x_2, \theta_2) - \gamma_1 \, x_1$$
$$\dot{x}_2 = \kappa_2 \, s^-(x_1, \theta_1) - \gamma_2 \, x_2$$



$$\dot{x} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{cases} \begin{bmatrix} \kappa_1 \\ \kappa_2 \end{bmatrix} - \begin{bmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, & \text{if } x \in \mathcal{X}_1 \\ \begin{bmatrix} \kappa_1 \\ 0 \end{bmatrix} - \begin{bmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, & \text{if } x \in \mathcal{X}_2 \\ \begin{bmatrix} 0 \\ \kappa_2 \end{bmatrix} - \begin{bmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, & \text{if } x \in \mathcal{X}_3 \\ - \begin{bmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, & \text{if } x \in \mathcal{X}_4 \end{cases}$$

Data model

Dynamics of the *i*-th molecule concentration:

$$\dot{x}_i(t) = \kappa_i^j - \gamma_i^j x_i(t) \text{ if } x(t) \in \mathcal{X}_j$$

Discrete-time model for the *i*-th molecule concentration:

$$x_i(k+1) = \tilde{\kappa}_i^j - \tilde{\gamma}_i^j x_i(k) + \eta_i(k) \text{ if } x(k) \in \mathcal{X}_j$$
$$y_i(k) = x_i(k) + \xi_i(k)$$

- T : sampling time step
- rate parameters: $\tilde{\kappa}_i^j = (\kappa_i^j/\gamma_i^j)(1 - e^{-\gamma_i^j T}) \quad ; \; \tilde{\gamma}_i^j = -e^{-\gamma_i^j T}$
- additive noise: η_i, ξ_i

Common data models:

- PieceWise Autoregressive eXogenous (PWARX): $\xi_i = 0$
- PWA Output-Error (PWA-OE): $\eta_i = 0$

PWA GRNs identification

Model:

$$\forall i \in \{1,\dots,n\}, \forall k \in \{1,\dots,N-1\}$$
$$x_i(k+1) = \tilde{\kappa}_i^j - \tilde{\gamma}_i^j x_i(k) + \eta_i(k) \text{ if } x(k) \in \mathcal{X}_j$$
$$y_i(k) = x_i(k) + \xi_i(k)$$

Identification problem: reconstruct

- the number of modes ← Given $x(0)$, how many $\mathcal{X}$ are reached ?
- all rate parameters ← For all of them, estimate $\tilde{\kappa}_i^j, \tilde{\gamma}_i^j$
- all switching thresholds ← Could some gene interactions cause the switch ?

from the noisy dataset $\mathcal{N} = \{y(k)\}_{k=0}^{N}$

Simple example

« biologocal » model ?

$$\dot{x}_1 = \kappa_1 s^-(x_2, \theta_2) - \gamma_1 x_1$$
$$\dot{x}_2 = \kappa_2 s^-(x_1, \theta_1) - \gamma_2 x_2$$
on $\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3, \mathcal{X}_4$

identifiable model

$$\dot{x}_1 = \kappa_1 s^-(x_2, \theta) - \gamma_1 x_1$$
$$\dot{x}_2 = \kappa_2 - \gamma_2 x_2$$
on $\mathcal{X}_1, \mathcal{X}_3$ given $x(0)$

features to identify

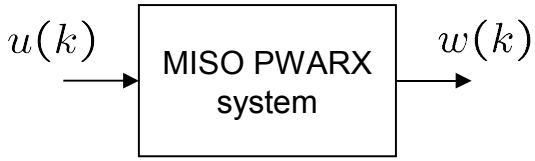
- 2 modes
- kinetic parameters: $\tilde{\kappa}^1, \tilde{\gamma}^1, \tilde{\kappa}^2, \tilde{\gamma}^2$
- switching threshold: $\tilde{\theta}$

Overview

- Genetic regulatory networks in brief
- Identification for PWA models of genetic regulatory networks
- PWA system identification
- Description of our approach
- Reconstruction of switching thresholds
- A case study: carbon starvation response of *E. coli*
- Conclusions

PWA identification methods

PWARX / PWA-OE models considered in hybrid identification:



$$z(k+1) = \pi^j [r(k)'1]' + \eta(k) \text{ if } r(k) \in \mathcal{X}_j$$
$$w(k) = z(k) + \xi(k)$$

$r(k) = [u'(k) \dots u'(k-n_a)\ z'(k) \dots z'(k-n_b)]'$ is the regressor vector.

$\{\mathcal{X}_j\}_{j=1}^{\tilde{s}}$ is a polyhedral partition of $\mathcal{X}$.

Dataset = noisy samples

$\mathcal{N} = \{ (r(k), y(k)) \}_{k=1}^{N}$

Common assumptions:

4. Known model orders
5. Known regressor set $\mathcal{X} \subset \Re^n$

Estimate:

2. The number $\tilde{s}$ of modes
3. The parameter vectors $\{\pi^j\}_{j=1}^{\tilde{s}}$
4. The regions $\{\mathcal{X}_j\}_{j=1}^{\tilde{s}}$

Identification in hybrid systems

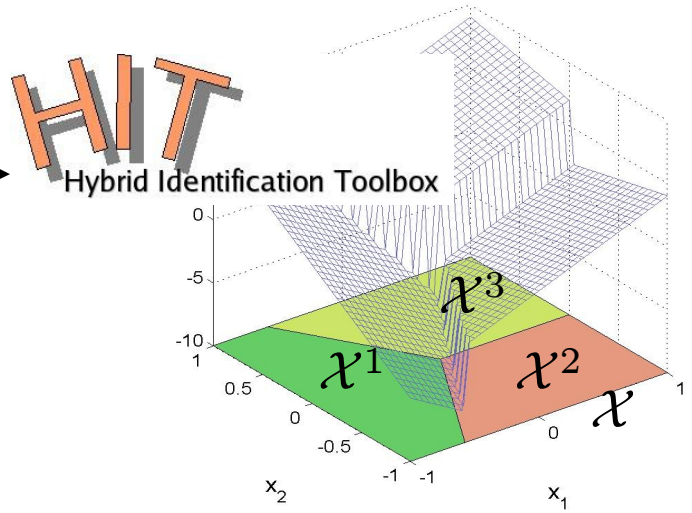
Some methods :

PWARX system identification:

- Ferrari-Trecate et al., 2003
- Bemporad et al., 2005
- Vidal et al., 2005
- Juloski et al., 2005

PWA-OE system identification:

- Juloski & Weiland, 2006
- Rosenqvist & Karlström, 2006

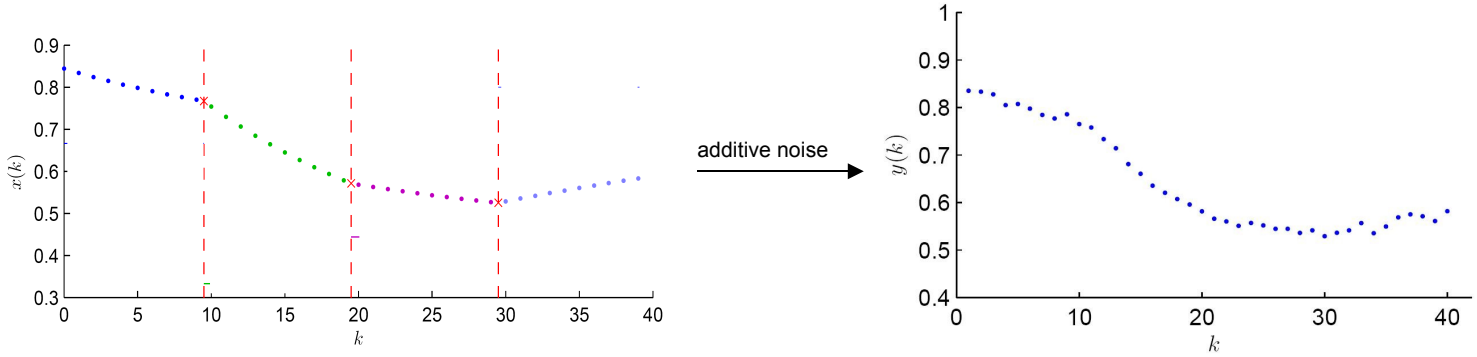


PWA models for a single molecule concentration fall within this class…

Hybrid identification pitfalls

- Existing identification methods are generic in nature and do not exploit features of PWA models of GRNs

Example 1: Switch detection from noisy measurements



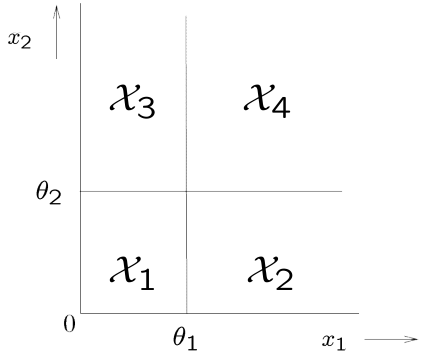
- Very challenging problem for general PWARX / PWA-OE models
- Much easier for PWA models of GRNs

Hybrid identification pitfalls

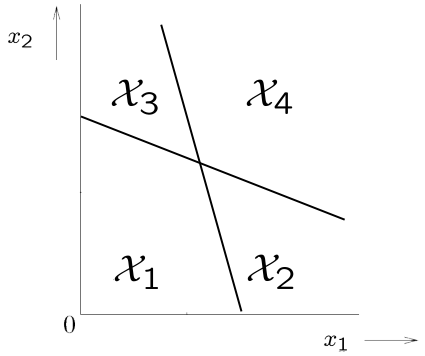
- Existing identification methods are generic in nature and do not exploit features of PWA models of GRNs
- Existing identification methods do not take into account constraints of PWA models of GRNs

Example 2: switching thresholds $\Rightarrow$ hyperrectangular domains

"true" domains



identified domains



The concept of threshold associated to a concentration variable is lost.

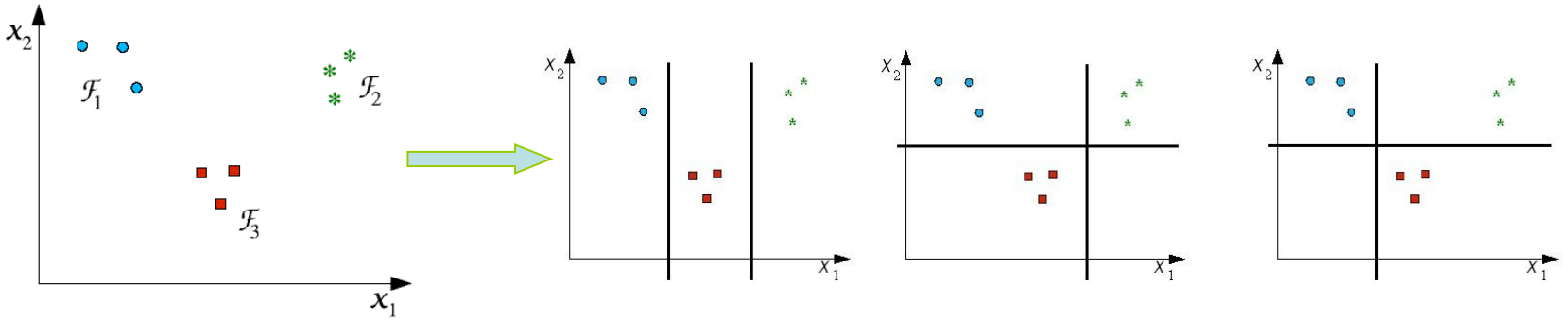
Hybrid identification pitfalls

- Existing identification methods are generic in nature and do not exploit features of PWA models of GRNs
- Existing identification methods do not take into account constraints of PWA models of GRNs
- Existing hybrid identification methods produce a single result but data are often scarce and multiple models might be plausible

Example 3: thresholds reconstruction

3 identified modes

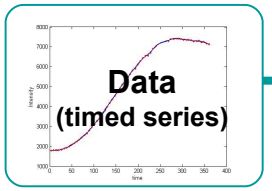
reconstructed thresholds: all sets are sufficient to explain the modes



Overview

- Genetic regulatory networks in brief
- Identification for PWA models of genetic regulatory networks
- PWA system identification
- Description of our approach
- Reconstruction of switching thresholds
- A case study: carbon starvation response of *E. coli*
- Conclusions

Identification of PWA models of GRNs



Switch detection

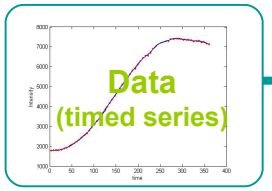
Mode estimation

Thresholds reconstruction

Kinetic parameters inference

Our approach:
« gray-box » identification.

Identification of PWA models of GRNs



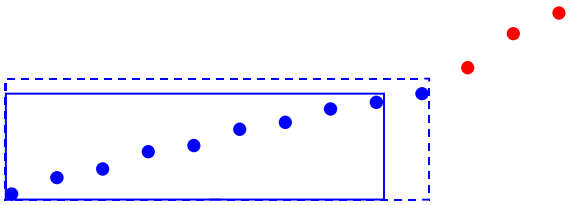
Switch detection

Mode estimation

Thresholds reconstruction

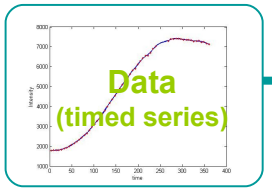
Kinetic parameters inference

1) Detection of switches in gene expression data



(Porreca et al., HSCC 2006)

Identification of PWA models of GRNs



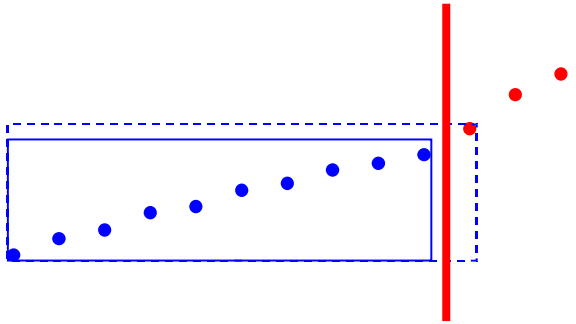
Switch detection

Mode estimation

Thresholds reconstruction

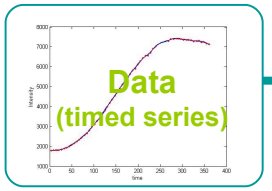
Kinetic parameters inference

1) Detection of switches in gene expression data



(Porreca et al., HSCC 2006)

Identification of PWA models of GRNs



Switch detection

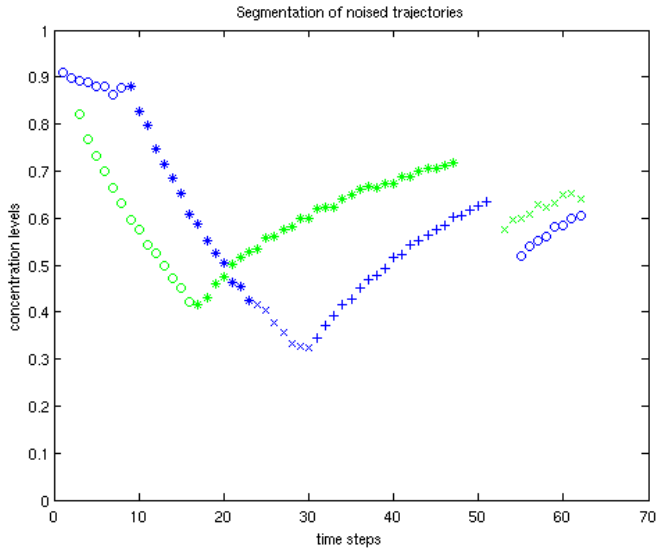
Mode estimation

Thresholds reconstruction

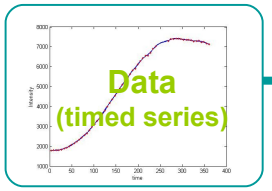
Kinetic parameters inference

2) Estimation of the number of modes and attribution of the measurements to mode data sets

Simple example



Identification of PWA models of GRNs



Switch detection

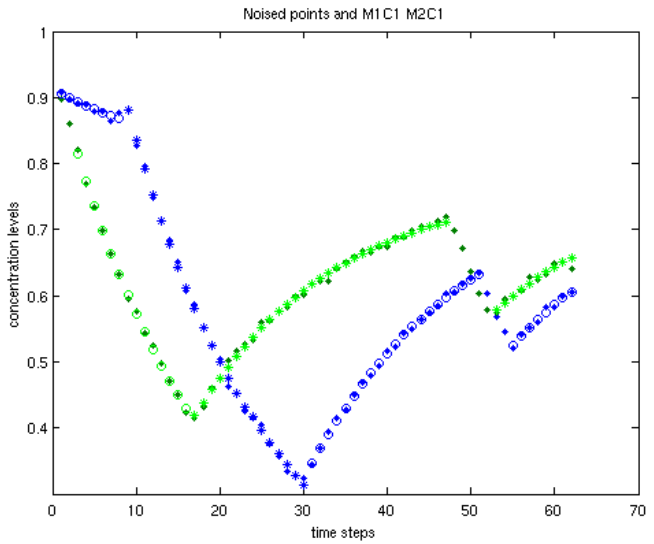
Mode estimation

Thresholds reconstruction

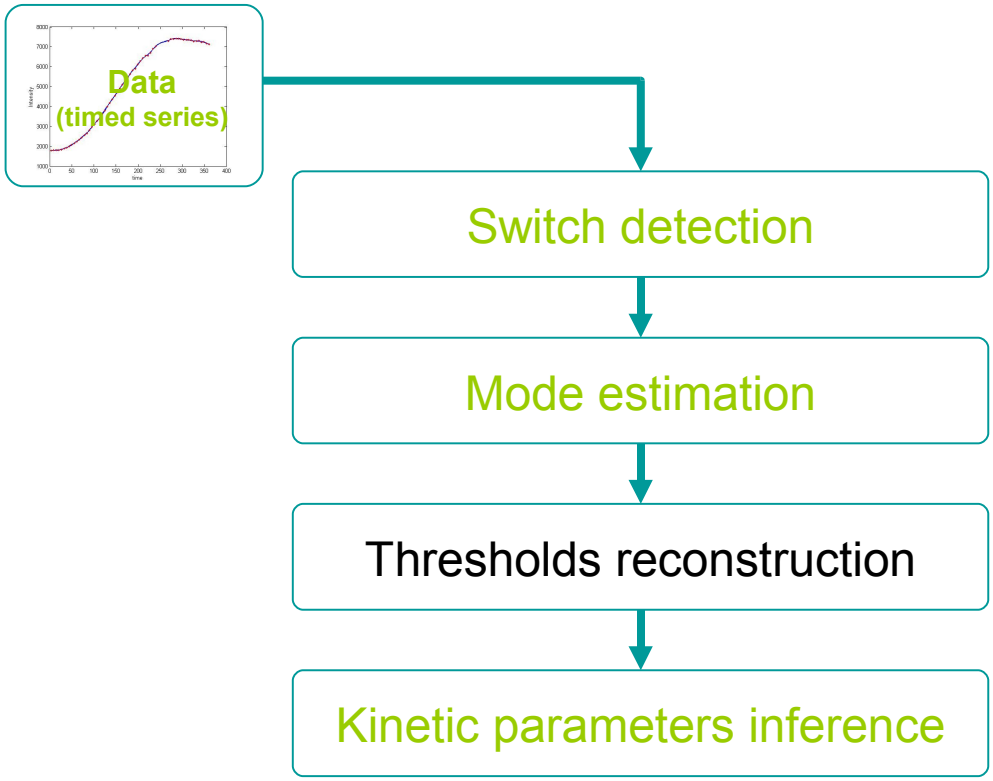
Kinetic parameters inference

2) Estimation of the number of modes and attribution of the measurements to mode data sets

Simple example



Identification of PWA models of GRNs

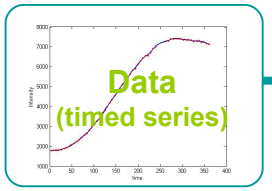


3) Reconstruction of

- thresholds on concentration variables
- all “minimal” combinations of thresholds consistent with the data

(Drulhe et al., HSCC 2005)

Identification of PWA models of GRNs



Switch detection

Mode estimation

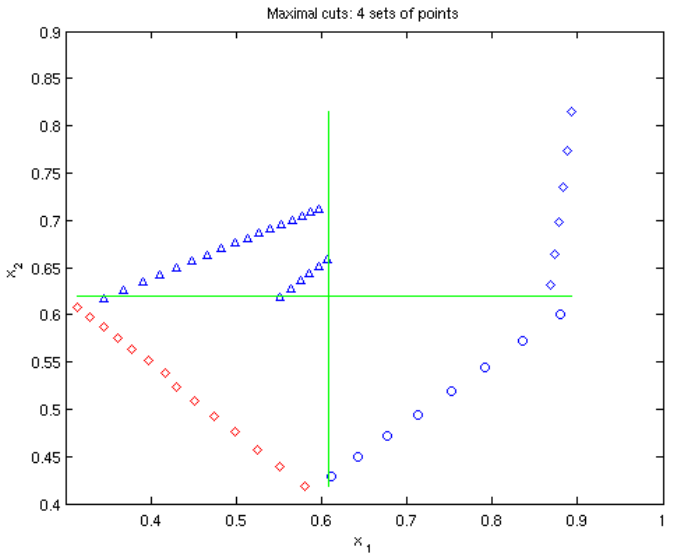
Thresholds reconstruction

Kinetic parameters inference

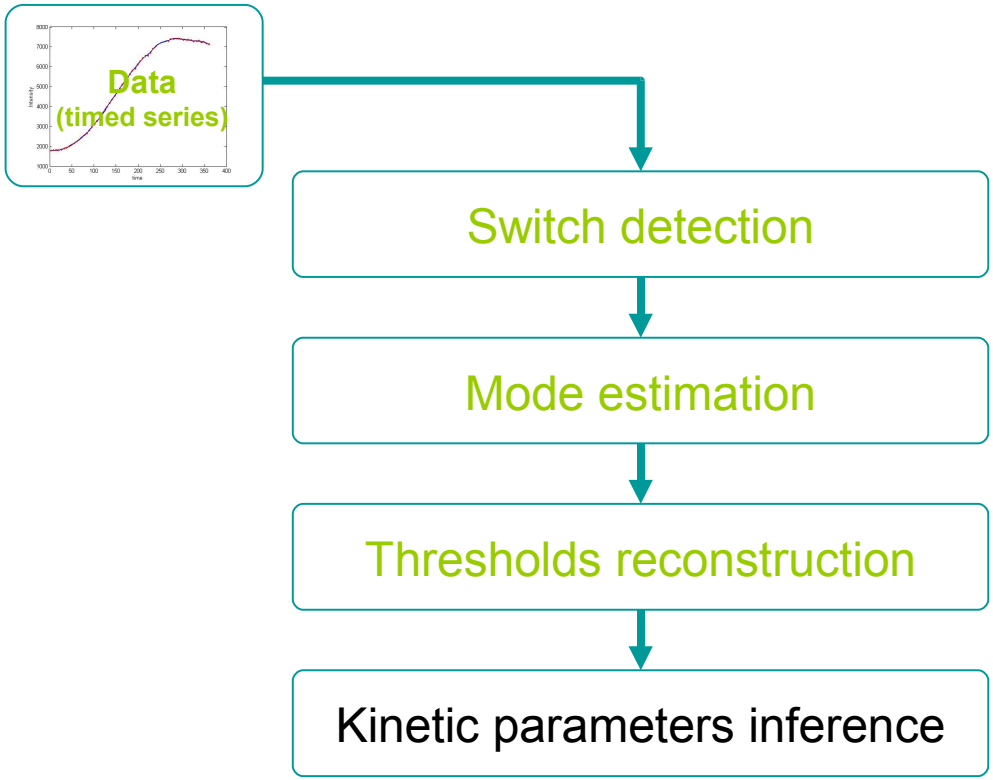
3) Reconstruction of

- thresholds on concentration variables

Simple example



Identification of PWA models of GRNs



4) Estimation of kinetic parameters for all models generated at previous stage

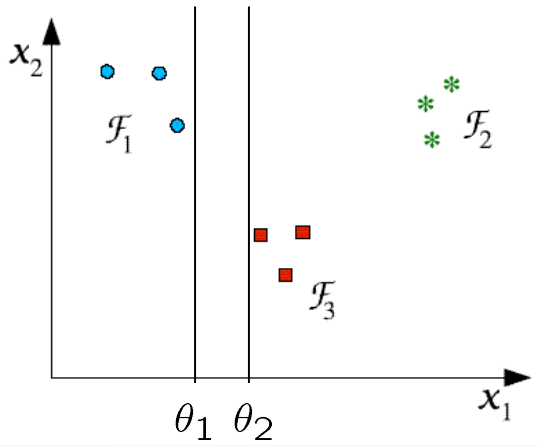
It is easy: LS on each mode data set.

Overview

- Genetic regulatory networks in brief
- Identification for PWA models of genetic regulatory networks
- PWA system identification
- Description of our approach
- Reconstruction of switching thresholds
- A case study: carbon starvation response of *E. coli*
- Conclusions

Switching thresholds as ap-hyperplanes

- An ap-hyperplane has a supporting vector parallel to one axis.
  - The label of the axis is the direction of the ap-hyperplane.
- The separation power $S(\theta)$ of an ap-hyperplane θ describes the separated data sets.
- Two ap-hyperplanes with a same direction and a same separation power are equivalent (thus defining equivalence classes of ap-hyperplanes).

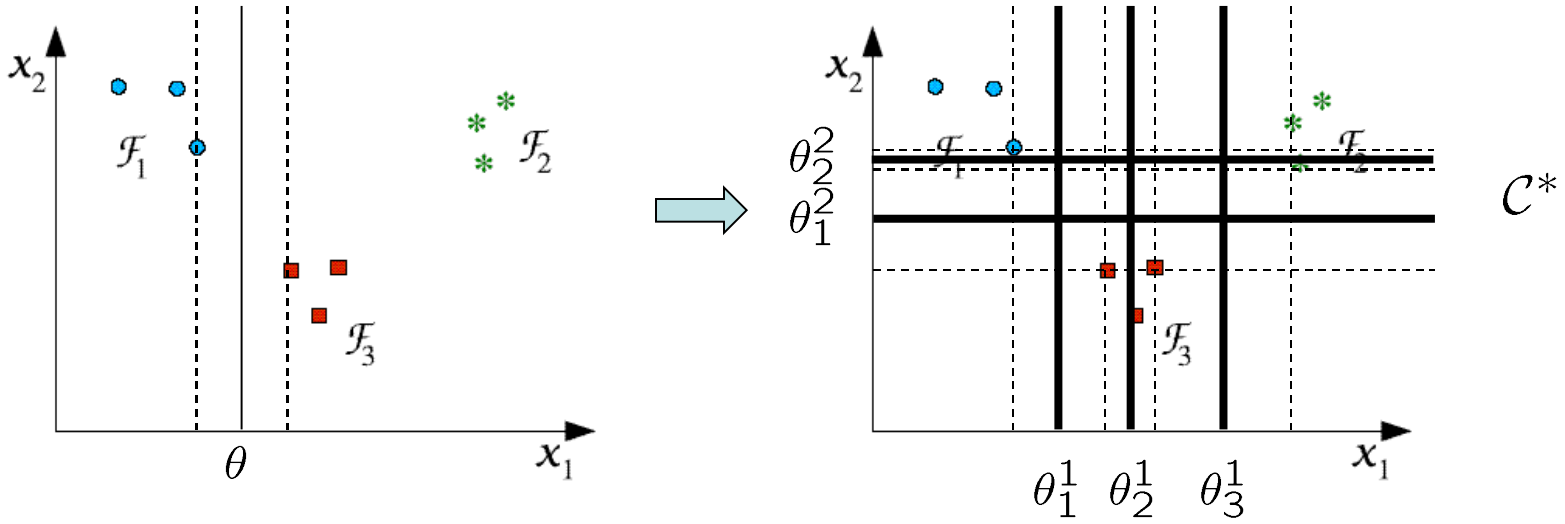


$$\begin{cases} S(\theta_1) = S(\theta_2) = \{(1,2),(1,3)\} \\ dir(\theta_1) = dir(\theta_2) = 1 \end{cases}$$

$$\Rightarrow \theta_1 \sim \theta_2$$

Switching thresholds and cuts

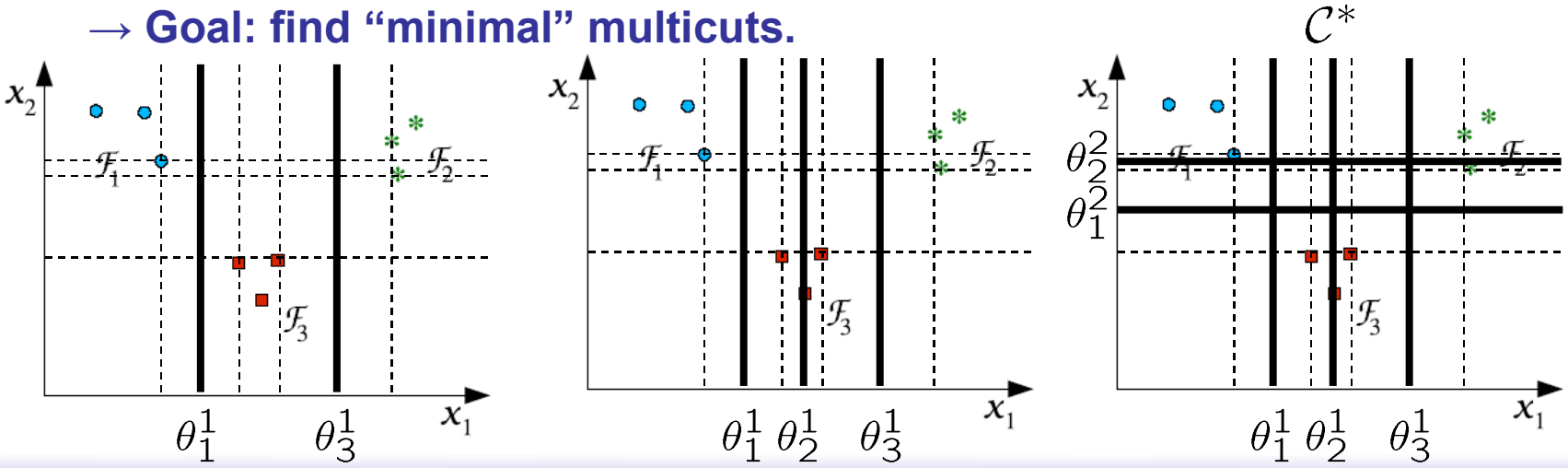
- For each class of equivalence, the ap-hyperplane that minimizes the empirical risk (i.e. that lies in the middle of the equivalence class) is a cut.
- The collection of all cuts $\mathcal{C}^*$ can be easily computed.
- Standing assumption: all pairs of sets are separated by at least one cut in $\mathcal{C}^*$



Switching thresholds and multicuts

- A collection of cuts such that any couple of data sets is separated by at least one of them is called a multicut.
- Any multicut is a collection of cuts that are sufficient to explain the observed dynamics. It provides a possible model of the genetic regulatory network.
- We are interested in finding models with a minimal number of interactions

→ **Goal: find "minimal" multicuts.**



Minimal multicuts

Assuming one can enumerate all possible multicuts from $\mathcal{C}^*$, a multicut is minimal if it has the *minimal cardinality*.

- Computation:
  - Rough idea: find all minimal multicuts by enumerating all multicuts ➔ combinatorial explosion!
  - Combinatorial optimization: one need a strategy to bound the search

- Mathematical ideas:
  - A cut is required if it is the only one to separate at least two mode data sets.
  - A cut is superfluous if all the required cuts separate at least all the same data sets as it does.

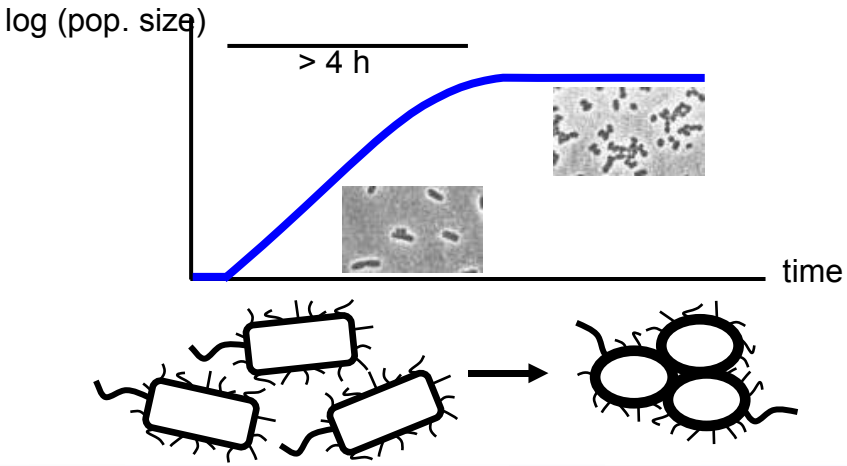
Overview

- Genetic regulatory networks in brief
- Identification for PWA models of genetic regulatory networks
- PWA system identification
- Description of our approach
- Reconstruction of switching thresholds
- A case study: carbon starvation response of *E. coli*
- Conclusions

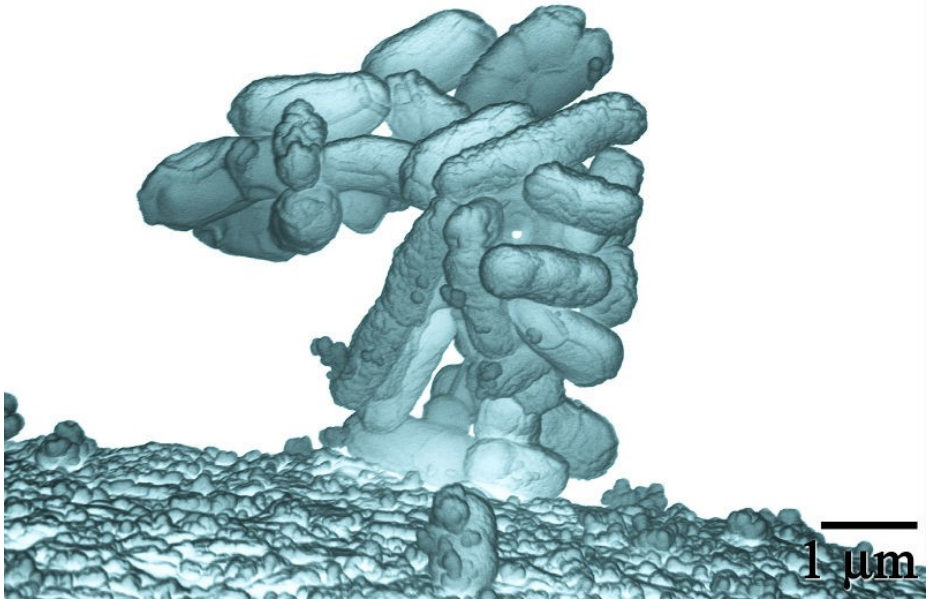
Application to *E.coli* carbon starvation response

Transitions from exponential to stationary phase involve observable changes in:

- morphology,
- metabolism,
- gene expression, …



Escherichia coli



Low-temperature electron micrograph of a cluster of E. coli bacteria, magnified 10,000 times. Photo by Eric Erbe, digital colorization by Christopher Pooley, both of USDA, ARS, EMU.

Application to *E.coli* carbon starvation response

Simplified PWA model of the response of *E. coli* to nutritional stress:

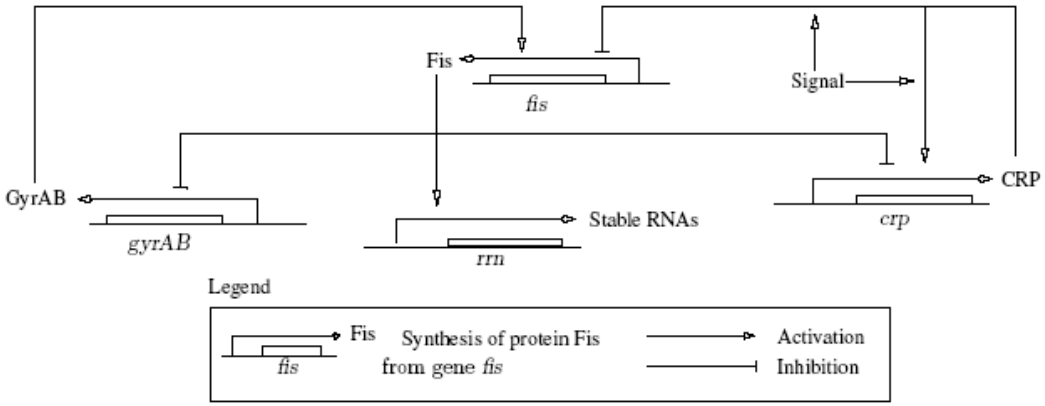
$$\dot{x}_{CRP} = \kappa^0_{CRP} + \kappa^1_{CRP}\, s^-(x_{Fis}, \theta^1_{Fis})\, s^+(x_{CRP}, \theta^1_{CRP})\, s^+(x_S, \theta_S) - \gamma_{CRP}\, x_{CRP}$$

$$\dot{x}_{Fis} = \kappa^1_{Fis}\,(1 - s^+(x_{CRP}, \theta^1_{CRP})\, s^+(x_S, \theta_S)) + \kappa^2_{Fis}\, s^+(x_{GyrAB}, \theta_{GyrAB})\,(1 - s^+(x_{CRP}, \theta^1_{CRP})\, s^+(x_S, \theta_S)) - \gamma_{Fis}\, x_{Fis}$$

$$\dot{x}_{GyrAB} = \kappa_{GyrAB}\, s^-(x_{Fis}, \theta^3_{Fis}) - \gamma_{GyrAB}\, x_{GyrAB}$$

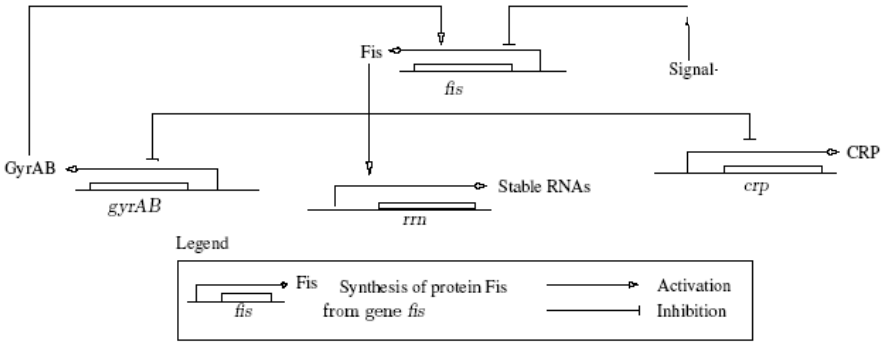
$$\dot{x}_{rrn} = \kappa_{rrn}\, s^+(x_{Fis}, \theta^2_{Fis}) - \gamma_{rrn}\, x_{rrn}$$

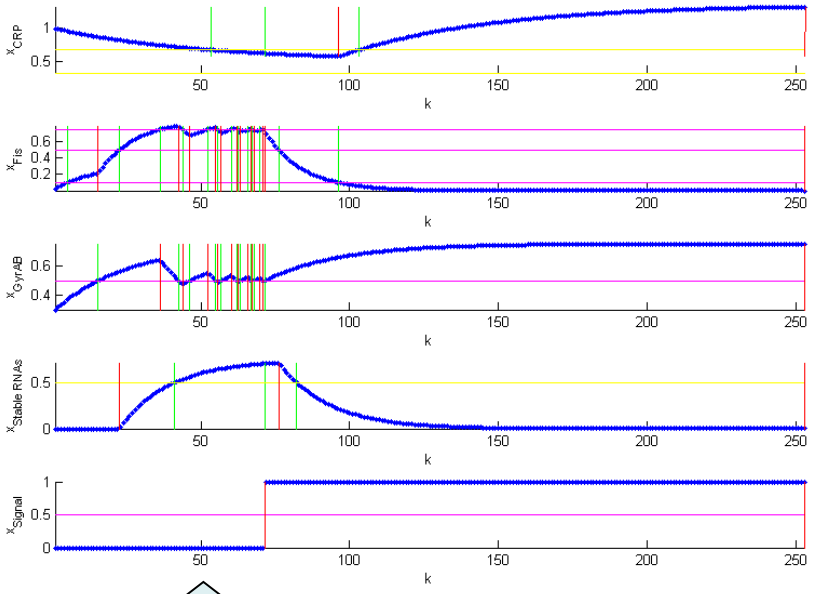
$$\dot{x}_S = 0$$



Application to *E.coli* carbon starvation response

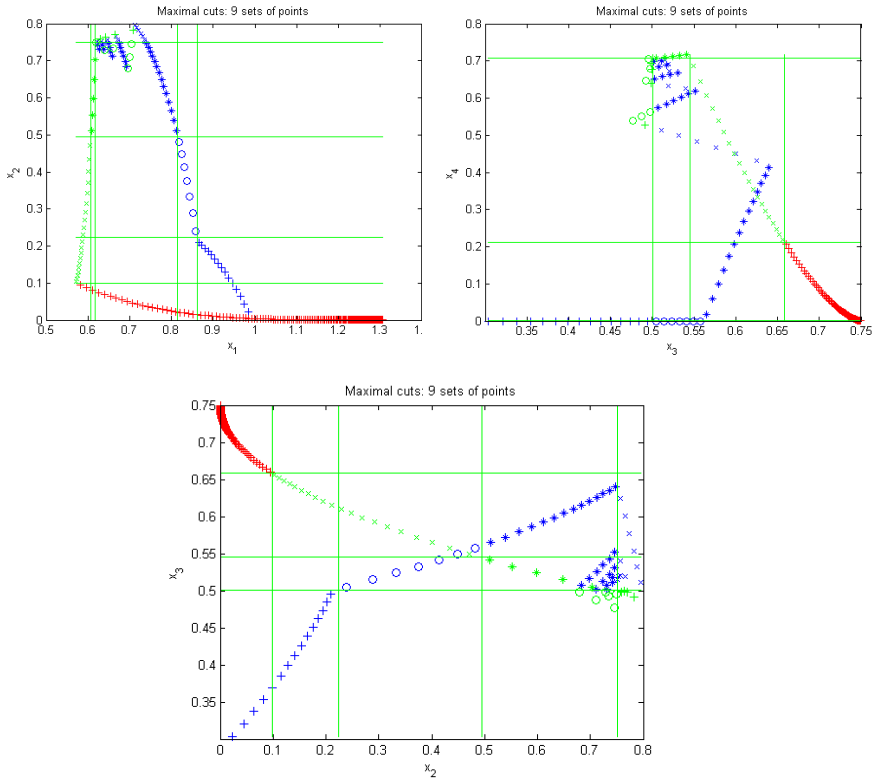
→ reconstruction of the interactions from simulated data.





Switching threshold reconstruction

The simulated data correspond to transitions exp→stat and stat→exp.



Cut	Variable	Threshold value	Interaction	Correct? (Y/N)
#1	CRP	0.60	activator of the synthesis of Stable RNAs	
#2	CRP	0.61	activator of the synthesis of Fis	
#3	CRP	0.81	inhibitor of the synthesis of Stable RNAs	
#4	CRP	0.86	inhibitor of the synthesis of Fis	
#5	Fis	0.10	inhibitor of the synthesis of CRP	Y
#6	Fis	0.22	activator of the synthesis of Fis	
#7	Fis	0.49	activator of the synthesis of Stable RNAs	Y
#8	Fis	0.75	inhibitor of the synthesis of GyrAB, activator and inhibitor of the synthesis of Fis	Y
#9	GyrAB	0.50	activator and inhibitor of the synthesis of Fis and GyrAB	Y
#10	GyrAB	0.54	inhibitor of the synthesis of Stable RNAs	
#11	GyrAB	0.65	activator of the synthesis of CRP	
#12	Stable RNAs	8.8e-6	activator of the synthesis of Stable RNAs	
#13	Stable RNAs	0.21	inhibitor of the synthesis of CRP	
#14	Stable RNAs	0.70	inhibitor of the synthesis of Fis, activator of the synthesis of Stable RNAs	
#15	Signal	0.50	Inhibitor of the synthesis of Fis	Y

Switching threshold reconstruction

The simulated data correspond to transitions exp→stat and stat→exp.

Multicut composed of cuts #:	Correct? (Y/N)
2, 5, 7, 8, 9	N, Y, Y, Y, Y
2, 7, 8, 9, 11	N, Y, Y, Y, N
5, 7, 8, 9, 15	Y, Y, Y, Y, Y
7, 8, 9, 11, 15	Y, Y, Y, N, Y
7, 8, 9, 13, 15	Y, Y, Y, N, Y

For the best globally minimal multicuts, the multicut approach has inferred all the identifiable interactions from the data.

Cut	Variable	Threshold value	Interaction	Correct? (Y/N)
#1	CRP	0.60	activator of the synthesis of Stable RNAs	
#2	CRP	0.61	activator of the synthesis of Fis	
#3	CRP	0.81	inhibitor of the synthesis of Stable RNAs	
#4	CRP	0.86	inhibitor of the synthesis of Fis	
#5	Fis	0.10	inhibitor of the synthesis of CRP	Y
#6	Fis	0.22	activator of the synthesis of Fis	
#7	Fis	0.49	activator of the synthesis of Stable RNAs	Y
#8	Fis	0.75	inhibitor of the synthesis of GyrAB, activator and inhibitor of the synthesis of Fis	Y
#9	GyrAB	0.50	activator and inhibitor of the synthesis of Fis and GyrAB	Y
#10	GyrAB	0.54	inhibitor of the synthesis of Stable RNAs	
#11	GyrAB	0.65	activator of the synthesis of CRP	
#12	Stable RNAs	8.8e-6	activator of the synthesis of Stable RNAs	
#13	Stable RNAs	0.21	inhibitor of the synthesis of CRP	
#14	Stable RNAs	0.70	inhibitor of the synthesis of Fis, activator of the synthesis of Stable RNAs	
#15	Signal	0.50	Inhibitor of the synthesis of Fis	Y

Conclusions

- Reconstruction of the switching thresholds consistent with the data
- Computation of different minimal models that are sufficient to explain the observed switches
- Example demonstrates applicability of this approach

Work in progress:

- Evaluation of the performance of the processing chain
- Validate the results on the full network of *E.Coli* with gene expression data (reporter genes)
- Optimization

Support:
European Commission under project HYGEIA (NEST-4995)