Models and data

Hidde de Jong
IBIS
INRIA Grenoble – Rhône-Alpes
Hidde.de-Jong@inria.fr

December 9, 2020

INRIA Grenoble - Rhône-Alpes and IBIS

- IBIS: systems biology group at INRIA/Université Joseph Fourier/CNRS
  - Analysis of bacterial regulatory networks by means of models and experiments
  - Biologists, computer scientists, mathematicians, physicists, …

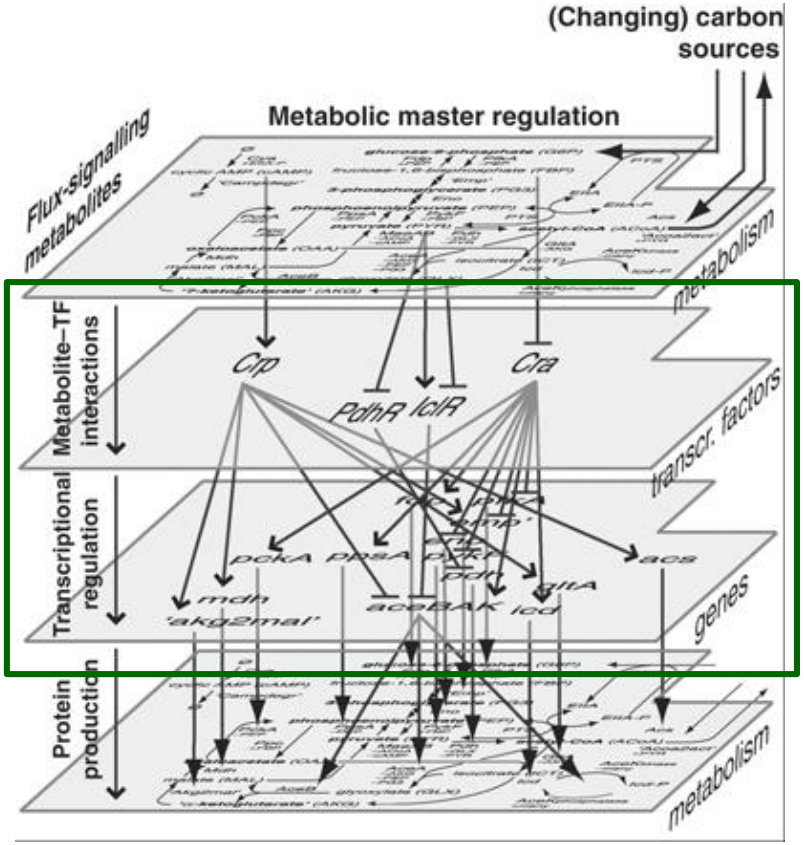
http://team.inria.fr/ibis

Overview

- Part 1. Systems biology and kinetic modeling
- Part 2. Metabolic network modeling
- Part 3. Gene regulatory network modeling
  - Quantitative modeling of gene regulatory networks
  - Qualitative modeling of gene regulatory networks
  - Stochastic modeling of gene regulatory networks
  - Practical on integrated models of bacterial growth (Matlab)
- **Part 4. Models and data**

Gene regulatory networks

- **Gene regulatory networks** control changes in gene expression levels in response to environmental perturbations



- Gene regulatory networks consist of genes, gene products, signalling metabolites, and their mutual regulatory interactions

  Global regulators of transcription involved in glucose-acetate diauxie in *E. coli*

Kotte *et al.* (2010), *Mol. Syst. Biol.,* 6:355

Modeling of gene regulatory networks

- Well-established theory for modeling of gene regulatory networks using ordinary differential equation (ODE) models

  Bolouri (2008), *Computational Modeling of Gene Regulatory Networks*, Imperial College Press

  Polynikis *et al.* (2009), *J. Theor. Biol.*, 261(4):511-30

- Practical problems encountered by modelers:
  - Knowledge on molecular mechanisms rare
  - **Quantitative information on kinetic parameters and molecular concentrations absent**
  - Large models

Lack of quantitative information: strategies

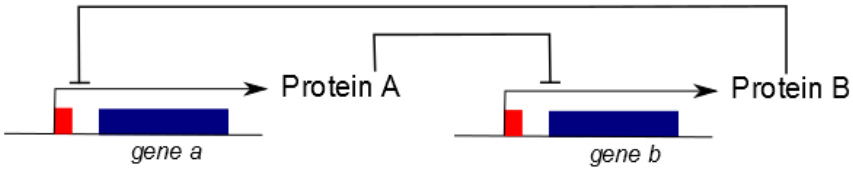
- Three main strategies to deal with lack of quantitative data:
  - Test of parameter sensitivity
  - Model reduction and simplification
  - Parameter estimation from time-series data

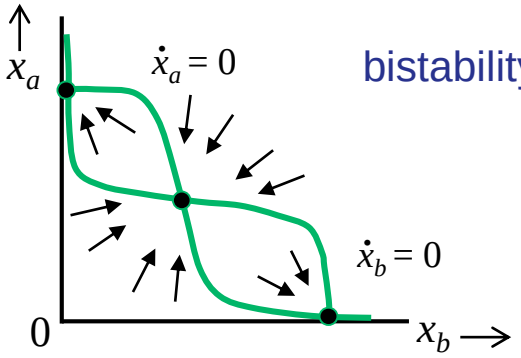
De Jong and Ropers (2006), *Brief. Bioinform.*, 7(4):354-363

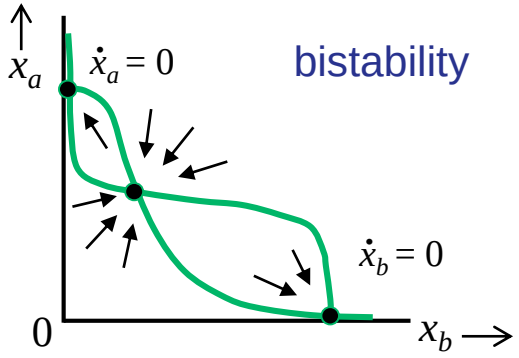
Test of parameter sensitivity

- Important dynamic properties are expected to be **robust** over large ranges of parameter values
  - Important dynamic properties should be insensitive to moderate variations in parameter values

Stelling *et al.* (2004), *Cell*, 118(6):675-685







Test of parameter sensitivity

- Important dynamic properties are expected to be **robust** over large ranges of parameter values

  Important dynamic properties should be insensitive to moderate variations in parameter values

  Stelling *et al.* (2004), *Cell*, 118(6):675-685

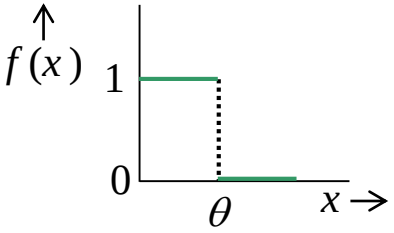
- Large variety of techniques for assessing sensitivity of models to changes in parameter values

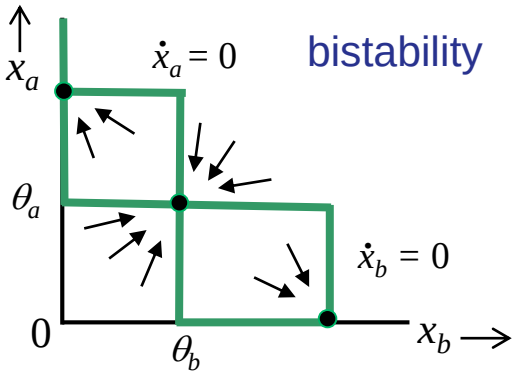
  Saltelli *et al.* (2008), *Global Sensitivity Analysis: The Primer. John Wiley & Sons.*

Model reduction and simplification

- Use model reduction and simplification to obtain models that can be analyzed with less information on parameter values
  - Piecewise-linear instead of nonlinear models
  - Also: Boolean models

$$\dot{x}_a = \kappa_a f(x_b) - \gamma_a x_a$$
$$\dot{x}_b = \kappa_b f(x_a) - \gamma_b x_b$$





Glass and Kauffman (1973), *J. Theor. Biol.*, 39(1):103-29

de Jong *et al.* (2004), *Bull. Math. Biol.*, 66(2):301-40

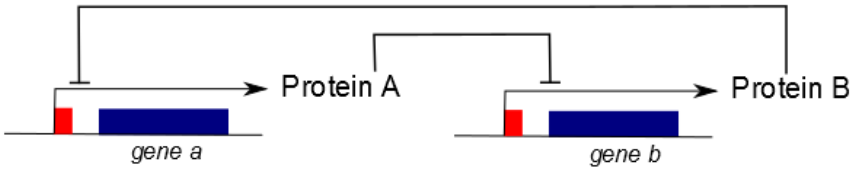
Parameter estimation

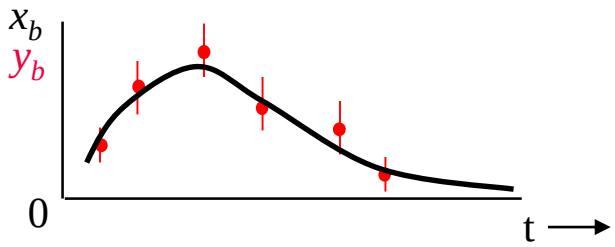
- **Estimate** parameter values from experimental time-series data

  Systems identification in control and engineering

  Walter and Pronzato (1997), *Identification of Parametric Models*, Springer

- Given model structure, search parameter values for which model predictions best fit experimental data





- Minimization of objective function, for instance sum of squared errors:

$$\sum_t (x(t,\theta) - y(t))^2$$

Parameter estimation problem

- Differential algebraic equation (DEA) models

$$\mathbf{A}\frac{d}{dt}\mathbf{x}(t,\mathbf{p}) = \mathbf{f}(\mathbf{x}(t,\mathbf{p}),\mathbf{u}(t),\mathbf{p}), \quad t_0 \leq t \leq t_e$$
$$\mathbf{x}(t_0,\mathbf{p}) = \mathbf{x}_0$$

  - Selection matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$
  - State variables $\mathbf{x} : \mathbb{R} \times \mathbb{R}^m \rightarrow \mathbb{R}^n$
  - Input variables $\mathbf{u} : \mathbb{R} \rightarrow \mathbb{R}^l$
  - Parameters $\mathbf{p} \in \mathbb{R}^m$
  - Rate functions $\mathbf{f} : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \rightarrow \mathbb{R}^n$

Ashyraliyev *et al.* (2009), *FEBS J,* 276(4):886–902

Parameter estimation problem

- Differential algebraic equation (DEA) models

$$\mathbf{A}\frac{d}{dt}\mathbf{x}(t,\mathbf{p}) = \mathbf{f}(\mathbf{x}(t,\mathbf{p}),\mathbf{u}(t),\mathbf{p}), \quad t_0 \leq t \leq t_e$$
$$\mathbf{x}(t_0,\mathbf{p}) = \mathbf{x}_0$$

- Observables: $\mathbf{g}(\mathbf{x}(t,\mathbf{p}),\mathbf{u}(t),\mathbf{p})$

$$\mathbf{g} \ : \ \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \rightarrow \mathbb{R}^q$$

- (Non)linear constraints: $\mathbf{c}(\mathbf{x}(t,\mathbf{p}),\mathbf{u}(t),\mathbf{p}) \geq 0$

Ashyraliyev *et al.* (2009), *FEBS J,* 276(4):886–902

Parameter estimation problem

- Model predictions at time t_i

$$\mathbf{g}^i = \mathbf{g}(\mathbf{x}(t_i, \mathbf{p}), \mathbf{u}(t_i), \mathbf{p})$$

- Measurements of observables at time t_i

$$y_j \sim \mathcal{F}(\mu_j, \sigma_j)$$
$$\mathbf{y}^i = \mathbf{y}(t_i)$$

- Model discrepancies

$$\mathbf{G} = [\mathbf{g}^1, \ldots, \mathbf{g}^N]$$
$$\mathbf{Y} = [\mathbf{y}^1, \ldots, \mathbf{y}^N]$$
$$\mathbf{E}(\mathbf{p}) = \mid \mathbf{G} - \mathbf{Y} \mid$$

Parameter estimation problem

- Objective function
  - Weighted sum of squared residuals

$$V(\mathbf{p}) = \sum_{i=1}^{N} \sum_{j=1}^{q} \frac{(g_j^i - y_j^i)^2}{(\sigma_j^i)^2} = \mathbf{E}(\mathbf{p})^T \, \mathbf{W} \, \mathbf{E}(\mathbf{p})$$

  - Other objective functions possible, more adapted to other measurement models or practical considerations!
- **Parameter estimation problem**:

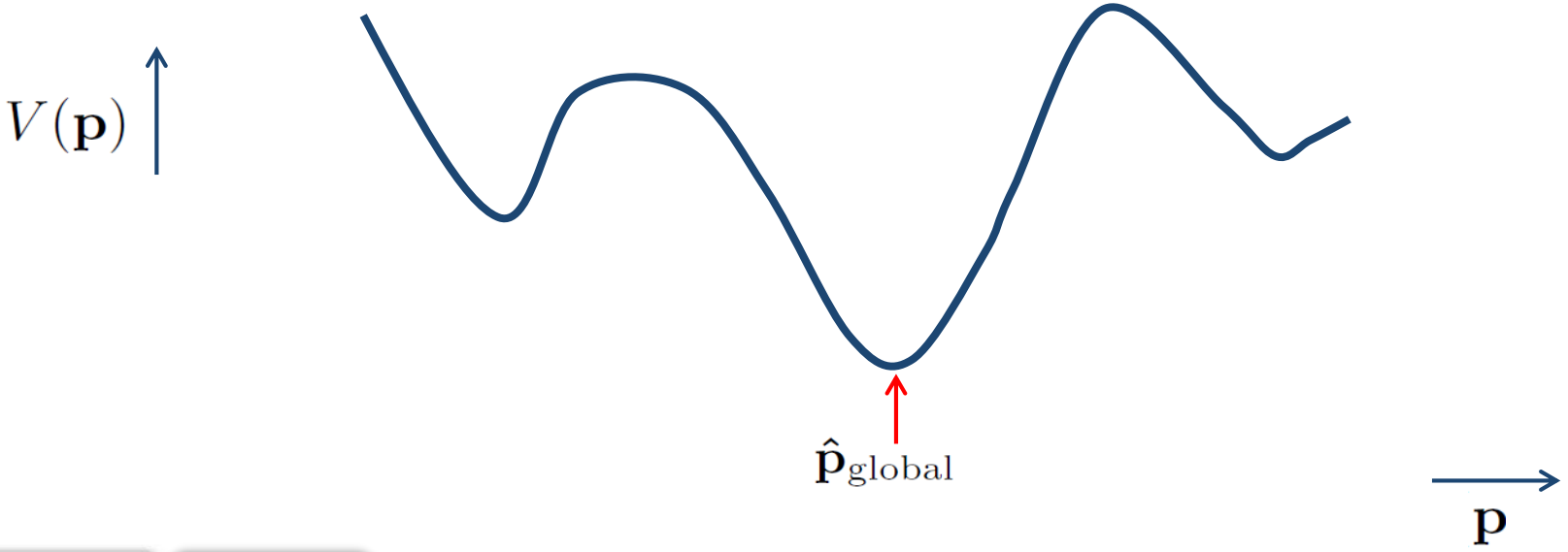
$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} V(\mathbf{p})$$

Parameter estimation methods

- Parameter estimation is a complex optimization problem
- Methods for solving optimization problem:
  - **Global methods**:

$$\hat{\mathbf{p}}_{\text{global}} = \arg\min_{\mathbf{p}} V(\mathbf{p})$$

for all $\mathbf{p}$ in parameter space

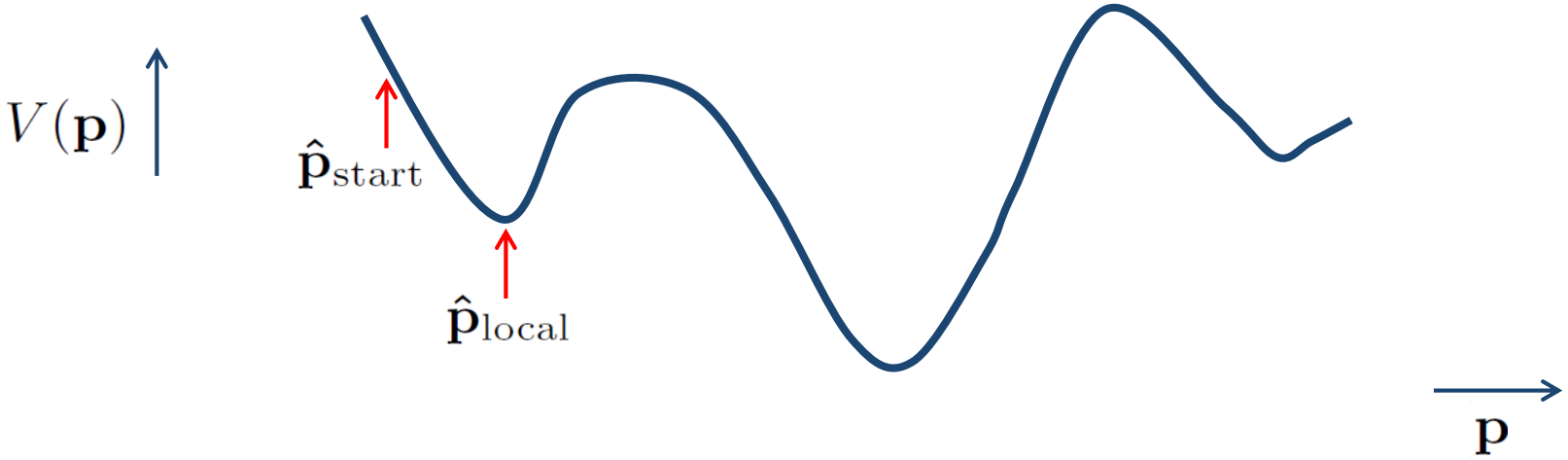


Parameter estimation methods

- Parameter estimation is a complex optimization problem
- Methods for solving optimization problem:
  - Global methods:
  - **Local methods**:

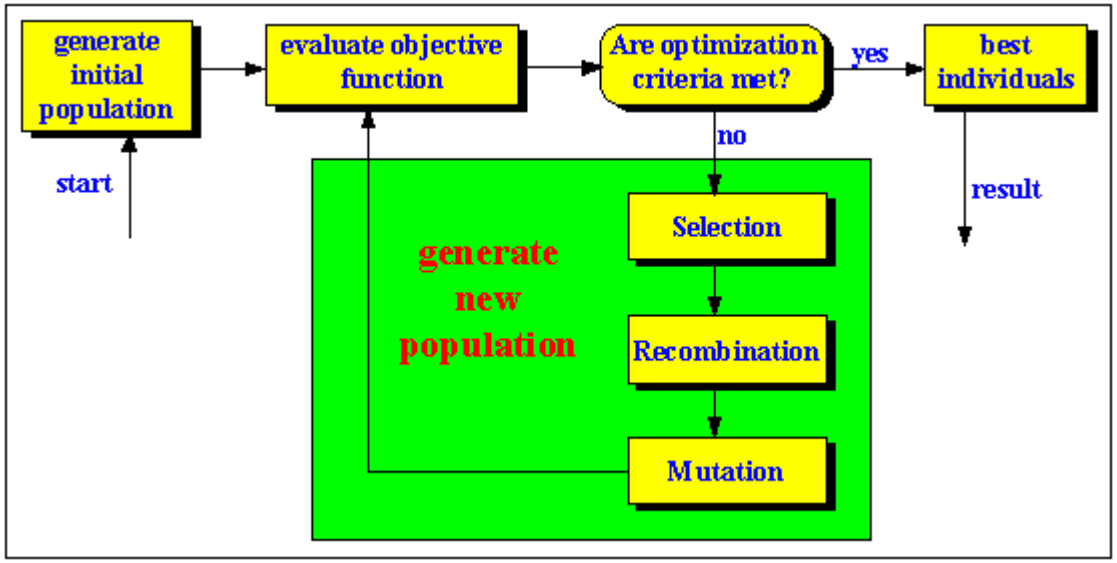
$$\hat{\mathbf{p}}_{\text{local}} = \arg\min_{\mathbf{p}} V(\mathbf{p})$$

for all $\mathbf{p}$ in a neighborhood of $\hat{\mathbf{p}}_{\text{start}}$



Global optimization methods

- Variety of global optimization methods:
  - Evolutionary algorithms



Ashyraliyev *et al.* (2009), *FEBS J,* 276(4):886–902

Global optimization methods

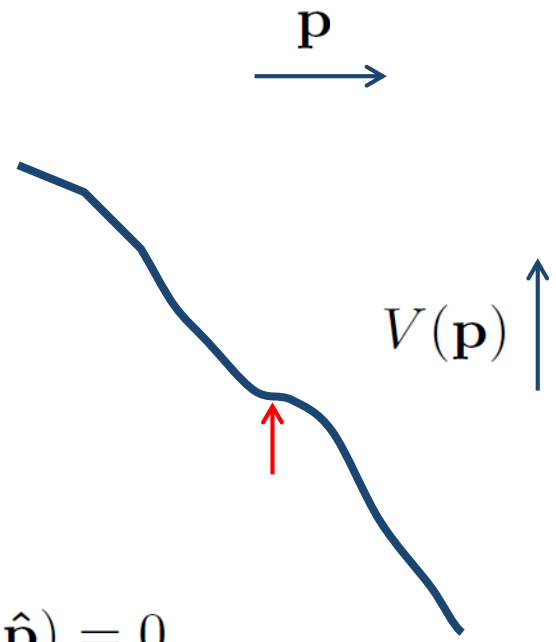
- Variety of global optimization methods:
  - Evolutionary algorithms
  - Simulated annealing
  - …
- Search of entire parameter space, but generally no convergence proof
- Mostly stochastic algorithms

Ashyraliyev *et al.* (2009), *FEBS J,* 276(4):886–902

Local optimization methods

- Variety of local optimization methods
  - Gradient methods

Gradient: $\nabla V(\mathbf{p}) = \left[\frac{\partial}{\partial p_i} V(\mathbf{p})\right]$



Necessary condition for local minimum: $\nabla V(\hat{\mathbf{p}}) = 0$

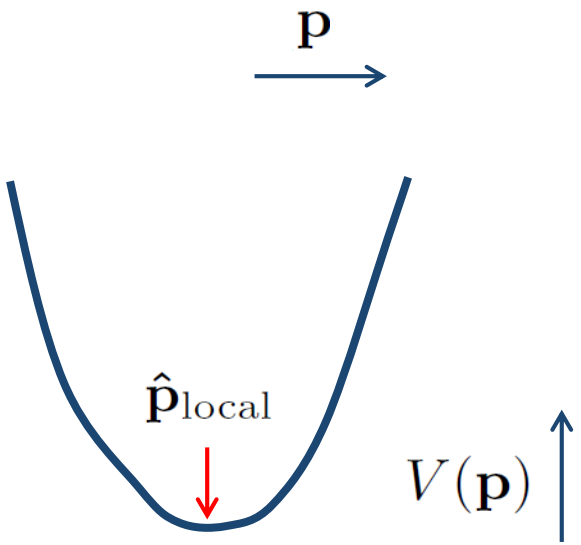
Ashyraliyev *et al.* (2009), *FEBS J,* 276(4):886–902

Local optimization methods

- Variety of local optimization methods
  - Gradient methods

Gradient: $\nabla V(\mathbf{p}) = \left[\frac{\partial}{\partial p_i} V(\mathbf{p})\right]$

Hessian: $\nabla^2 V(\mathbf{p}) = \left[\frac{\partial}{\partial p_i \partial p_j} V(\mathbf{p})\right]$



Necessary condition for local minimum: $\nabla V(\hat{\mathbf{p}}) = 0$

Sufficient condition for local minimum:

$$\nabla V(\hat{\mathbf{p}}) = 0 \quad \text{and} \quad \mathbf{p}^T \nabla^2 V(\hat{\mathbf{p}})\, \mathbf{p} > 0 \quad \text{for all} \quad \mathbf{p}$$

Local optimization methods

- Variety of local optimization methods
  - Gradient methods

$$\mathbf{p}_{\text{new}} = \mathbf{p} + \alpha\, d\mathbf{p}, \quad V(\mathbf{p}_{\text{new}}) < V(\mathbf{p})$$

Steepest descent

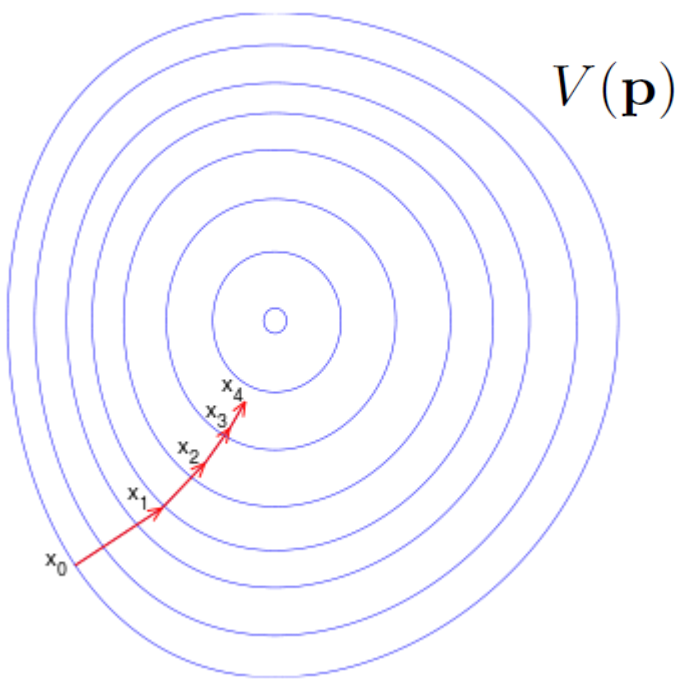
$$d\mathbf{p} = -\nabla V(\mathbf{p})$$

Newton's method

$$d\mathbf{p} = -\nabla^{-2} V(\mathbf{p})\, \nabla V(\mathbf{p})$$

Other choices for $d\mathbf{p}$

Adaptive choice of α



Local optimization methods

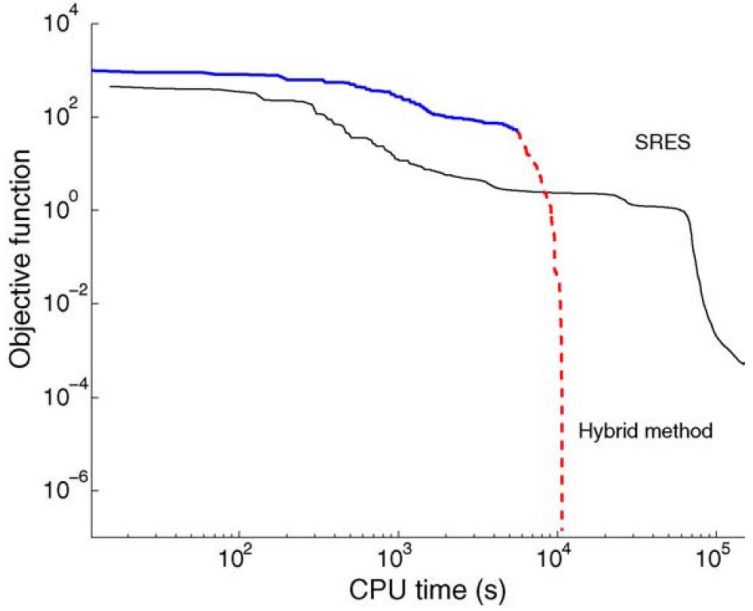
- Variety of local optimization methods
  - Gradient methods
  - Direct search
  - …
- Local search of parameter space, but proof of (speed of) convergence
- Mostly deterministic algorithms

Ashyraliyev *et al.* (2009), *FEBS J,* 276(4):886–902

Hybrid optimization method

- **Hybrid optimization**: global followed by local optimization
- In practice, hybrid optimization methods work well on large nonlinear models used in systems biology

  Test on benchmark identification problems



Rodriquez-Fernandez *et al.* (2006), *Biosystems*, 83: 248–65

Constrained optimization methods

- Differential algebraic equation (DEA) models

$$\mathbf{A}\frac{d}{dt}\mathbf{x}(t,\mathbf{p}) = \mathbf{f}(\mathbf{x}(t,\mathbf{p}),\mathbf{u}(t),\mathbf{p}), \quad t_0 \le t \le t_e$$
$$\mathbf{x}(t_0,\mathbf{p}) = \mathbf{x}_0$$

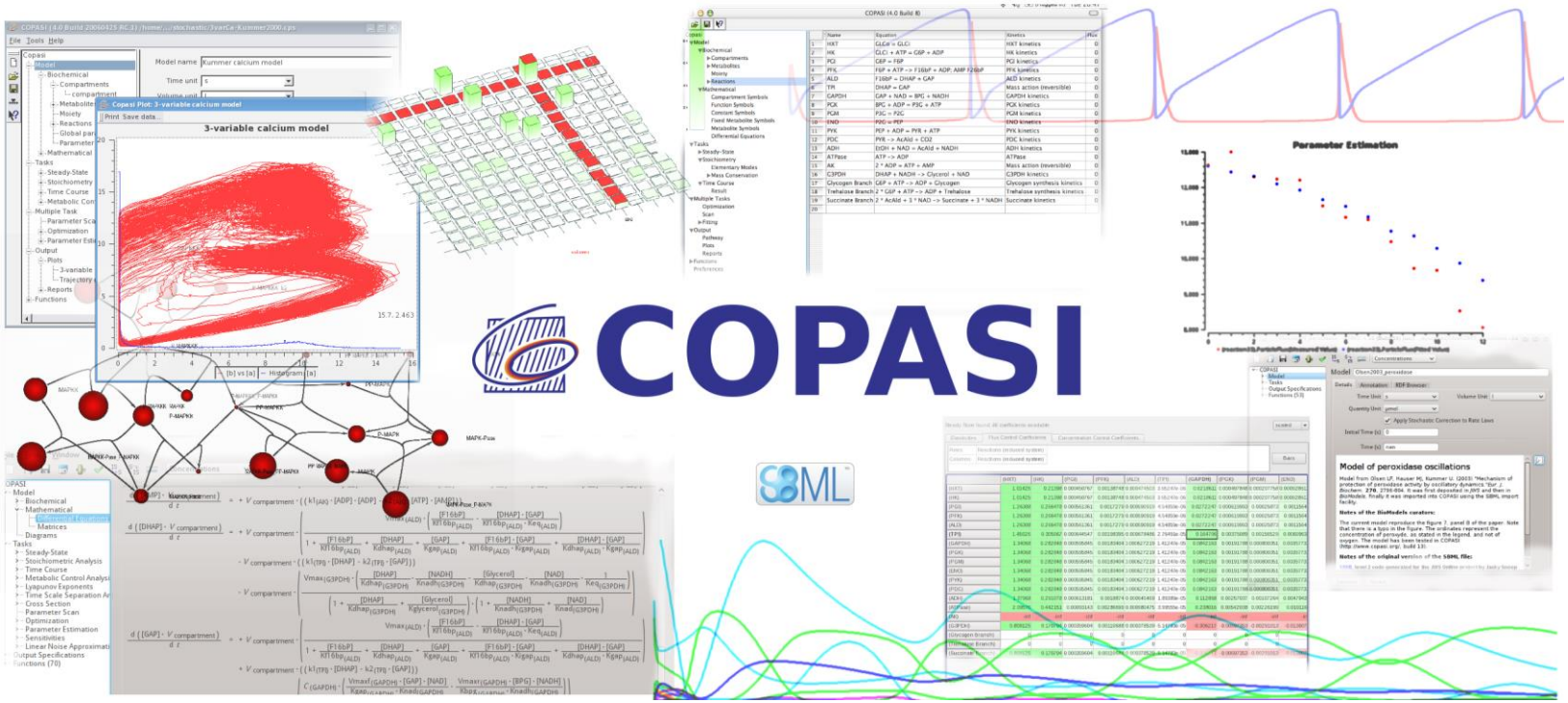
- (Non)linear constraints:

$$\mathbf{c}(\mathbf{x}(t,\mathbf{p}),\mathbf{u}(t),\mathbf{p}) \ge 0$$

- Constraints can be implemented in different ways
  - Penalty term in objective function
  - Search in subspace defined by (in)equality constraints
  - ….

Parameter estimation tools

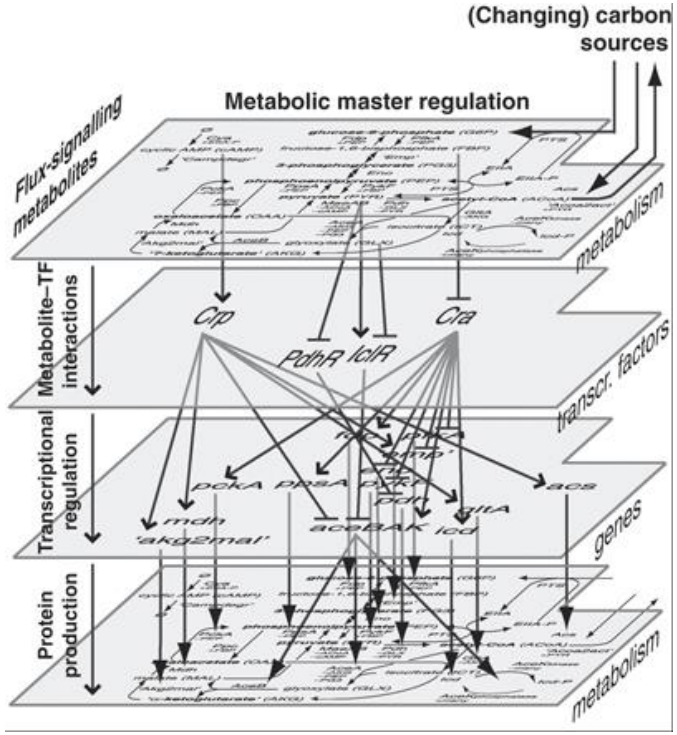
- Large number of dedicated parameter estimation tools in systems biology, in addition to general-purpose tools

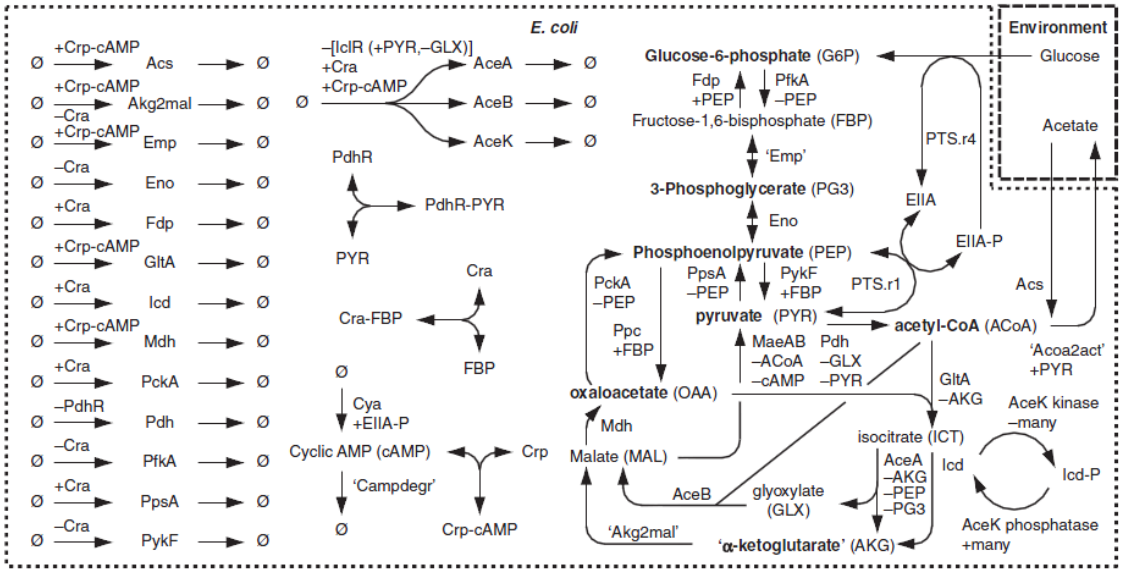


Hoops et al. (2006), *Bioinformatics,* 22:3067-74

Integrated model of *E. coli* metabolism

- Coupling of gene expression and metabolism into a single integrated model of *E. coli* metabolism
  - Kinetic model with 47 variables and 193 parameters





Kotte *et al.* (2010), *Mol. Syst. Biol.*, 6: 355

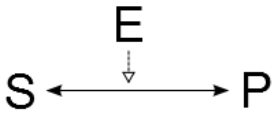
Multi-scale network of *E. coli* metabolism

- Estimation of model parameters using steady-state data: algebraic equations

  Published data sets for balanced growth on either glucose or acetate

$$N \cdot v = 0$$

- **Divide-and-conquer strategy** based on model structure
  1. Estimate reaction rates at steady state
  2. Identify model parameters for individual reactions from reaction rate estimations and measurements of concentrations



$$v(s,p) = \frac{V_m^+ \cdot s/K_{m1} - V_m^- \cdot p/K_{m2}}{1 + s/K_{m1} + p/K_{m2}}$$

Kotte *et al.* (2009), *Bioinformatics*, 25(4):519-25

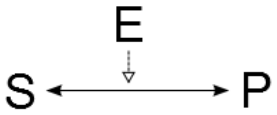
Multi-scale network of *E. coli* metabolism

- Estimation of model parameters using steady-state data: algebraic equations

  Published data sets for balanced growth on either glucose or acetate

$$N \cdot v = 0$$

- **Divide-and-conquer strategy** based on model structure
  1. Estimate reaction rates at steady state
  2. Identify model parameters for individual reactions from reaction rate estimations and measurements of concentrations



$$v(s,p) = \frac{V_m^+ \cdot s/K_{m1} - V_m^- \cdot p/K_{m2}}{1 + s/K_{m1} + p/K_{m2}}$$

Kotte *et al.* (2009), *Bioinformatics*, 25(4):519-25

Multi-scale network of *E. coli* metabolism

- Estimation of model parameters using steady-state data: algebraic equations

  Published data sets for balanced growth on either glucose or acetate

$$N \cdot v = 0$$

- **Divide-and-conquer strategy** based on model structure
  1. Estimate reaction rates at steady state
  2. Identify model parameters for individual reactions from reaction rate estimations and measurements of concentrations
  3. Re-estimate integrated model

$$f_{E,Pdh} = \frac{x_{Pdh}\, p_{Pdh,k_{cat}} \frac{x_{PYR}}{p_{Pdh,K_{PYR}}} \left(1 + \frac{x_{PYR}}{p_{Pdh,K_{PYR}}}\right)^{p_{Pdh,n}-1}}{\left(1 + \frac{x_{PYR}}{p_{Pdh,K_{PYR}}}\right)^{p_{Pdh,n}} + p_{Pdh,L}\left(1 + \frac{x_{GLX}}{p_{Pdh,K_{GLX}}} + \frac{x_{PYR}}{p_{Pdh,K_{I,PYR}}}\right)^{p_{Pdh,n}}}$$

$$f_{E,PfkA} = \frac{x_{PfkA}\, p_{PfkA,k_{cat}} \frac{x_{G6P}}{p_{PfkA,K_{G6P}}} \left(1 + \frac{x_{G6P}}{p_{PfkA,K_{G6P}}}\right)^{p_{PfkA,n}-1}}{\left(1 + \frac{x_{G6P}}{p_{PfkA,K_{G6P}}}\right)^{p_{PfkA,n}} + p_{PfkA,L}\left(1 + \frac{x_{PEP}}{p_{PfkA,K_{PEP}}}\right)^{p_{PfkA,n}}}$$

Kotte *et al.* (2010), *Mol. Syst. Biol.*, 6: 355

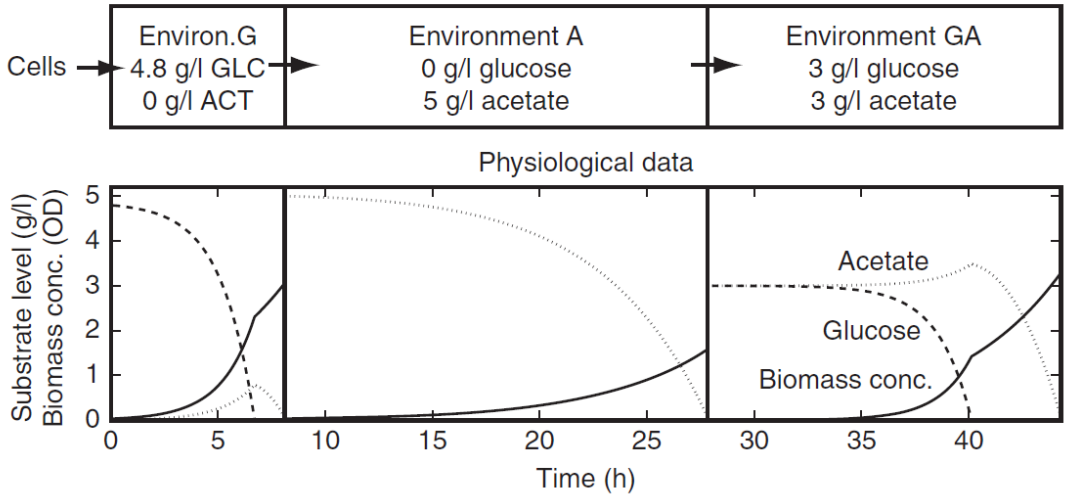
Multi-scale network of *E. coli* metabolism

- Estimation of model parameters using steady-state data: algebraic equations

  Published data sets for balanced growth on either glucose or acetate

$$N \cdot v = 0$$

- Model reproduces known physiological behavior of *E. coli*: diauxic growth and carbon catabolite repression



Kotte *et al.* (2010), *Mol. Syst. Biol.*, 6: 355

Identifiability analysis

- Parameter p_j is (globally) **identifiable**, if it can be uniquely determined from given model input $\mathbf{u}(t)$ and error-free data
- Model is (globally) identifiable, if all of its parameters are
- **A-priori identifiability** analysis

  Detect structural problems of model

$$\frac{d}{dt}x(t,\mathbf{p}) = p_1 \cdot x + p_2 \cdot x$$

  Given measurements of x, it is not possible to obtain independent estimates of both parameters

Walter and Pronzato (1997), *Identification of Parametric Models*, Springer

Identifiability analysis

- Parameter p_j is (globally) **identifiable**, if it can be uniquely determined from given model input $\mathbf{u}(t)$ and error-free data
- Model is (globally) identifiable, if all of its parameters are
- **A-priori identifiability** analysis

  Detect structural problems of model
- **Practical** or **a-posteriori identifiability** analysis
  - Problems with precision and quality of data
  - Correlation between variables
- Parameter p_j is practically identifiable if its confidence interval is of finite size
- Model is practically identifiable, if all of its parameters are

Identifiability analysis using profile likelihood

- Objective function

$$V(\mathbf{p}) = \sum_{i=1}^{N} \sum_{j=1}^{q} \frac{(g_j^i - y_j^i)^2}{(\sigma_j^i)^2} = \mathbf{E}(\mathbf{p})^T \, \mathbf{W} \, \mathbf{E}(\mathbf{p})$$

$$y_j \sim \mathcal{N}(\mu_j, \sigma_j) \text{ iid}$$

  is proportional to **log-likelihood**

  Minimization of $V(\mathbf{p})$ corresponds to maximum likelihood estimation

- Profile likelihood for parameter p_i ranging over interval $p_i = \tilde{p}_i \in [\underline{p}, \overline{p}]$ is defined as

$$V_{PL}(\tilde{p}_i) = \min_{\mathbf{p} \in \mathbb{R}^m,\ p_i = \tilde{p}_i} V(\mathbf{p})$$

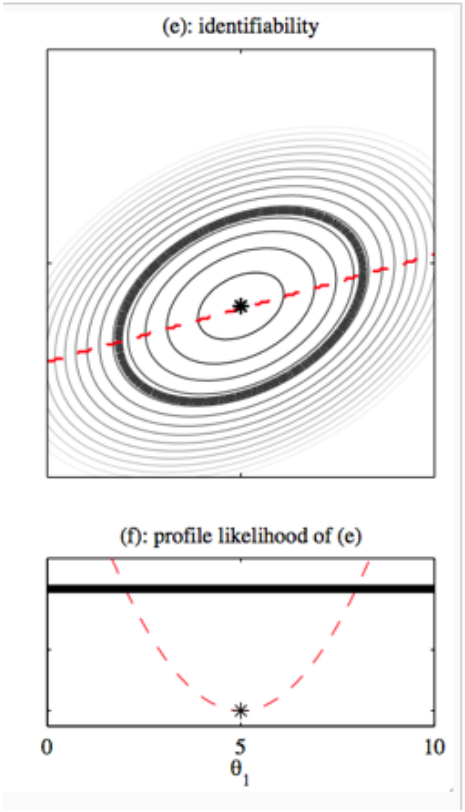
- Likelihood-based confidence interval $\{\mathbf{p} \mid V(\mathbf{p}) - V(\hat{\mathbf{p}}) < \alpha\}$

  Threshold given by distribution

Raue *et al.* (2009), *BIoinformatics*, 25(15):1923-9

Identifiability analysis using profile likelihood

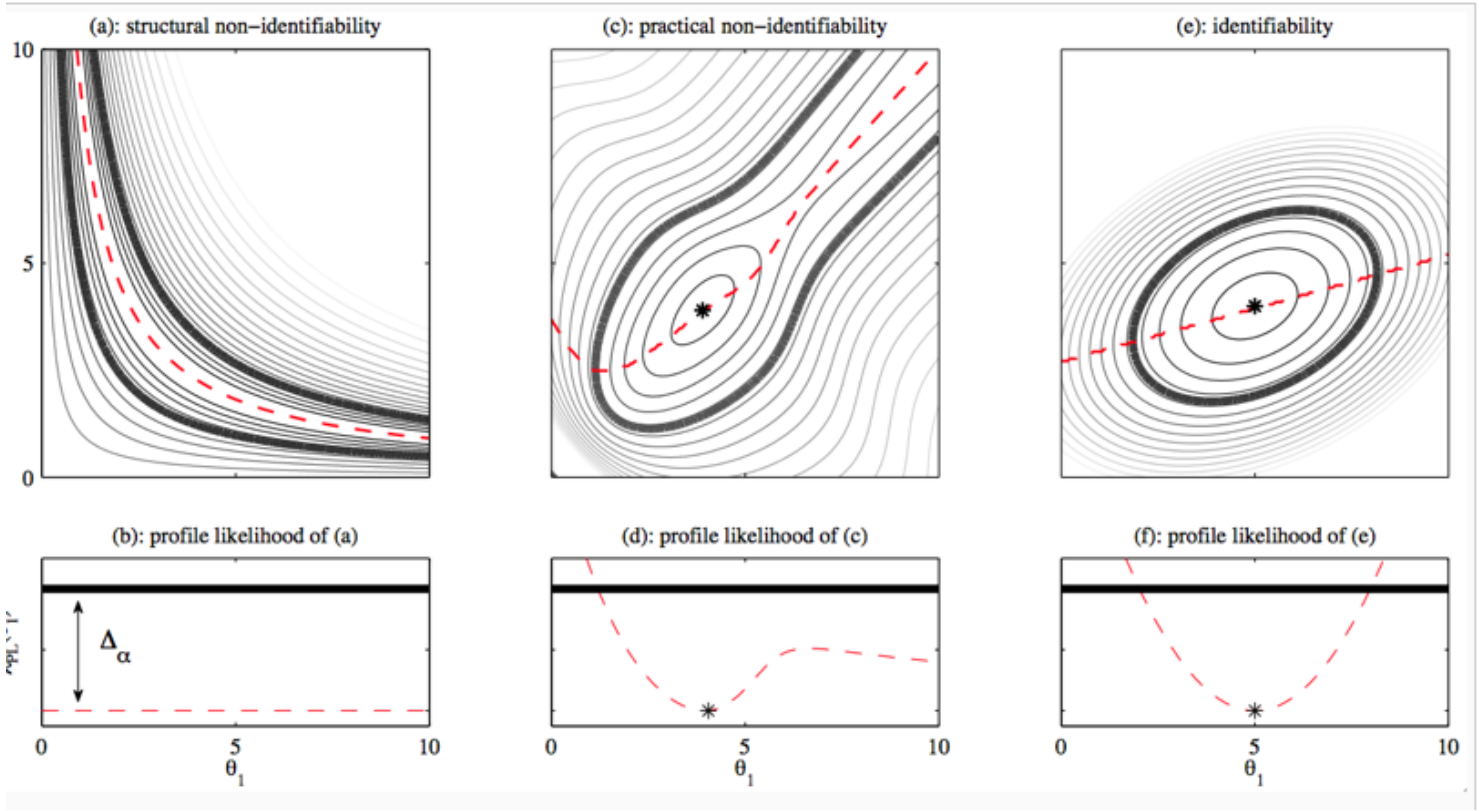
- Intuition: for each parameter, explore the parameter space in direction of least increase of objective function



Raue *et al.* (2011), *IET Syst. Biol.*, 5(2):120-30

Identifiability analysis using profile likelihood

- Intuition: for each parameter, explore the parameter space in direction of least increase of objective function



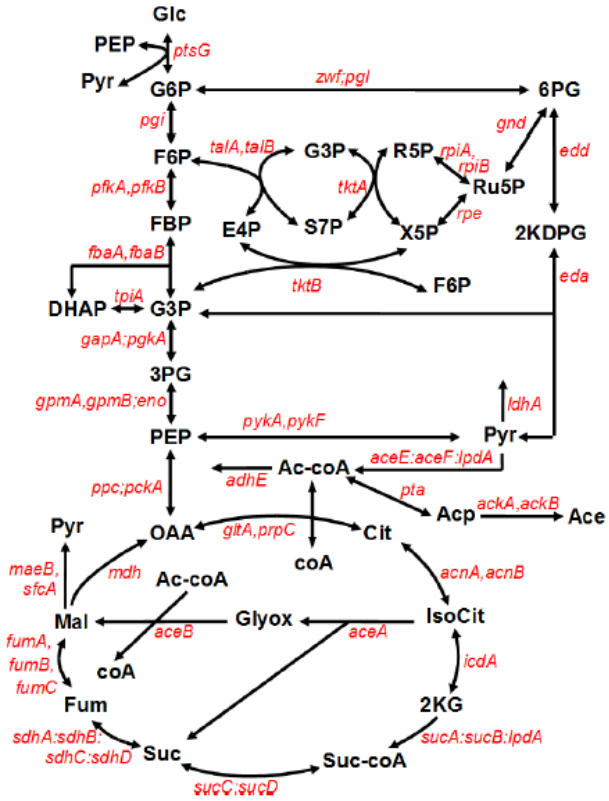
Raue *et al.* (2011), *IET Syst. Biol.*, 5(2):120-30

Identifiability analysis of *E. coli* metabolism

- Simplified model of central carbon metabolism in E. coli
- Model at steady state and lin-log kinetics for describing reactions

$$v = \mathrm{diag}(e) \cdot (a + B^x \cdot \ln x + B^u \cdot \ln u)$$
$$N \cdot v = 0$$

- Measurements:
  - Metabolite concentrations u, x
  - Enzyme concentrations e
  - Metabolic fluxes (reaction rates at steady state) J
- Parameters to estimate: a B^x B^u



Berthoumieux *et al.* (2011), *Bioinformatics*, 27(13):i186–93

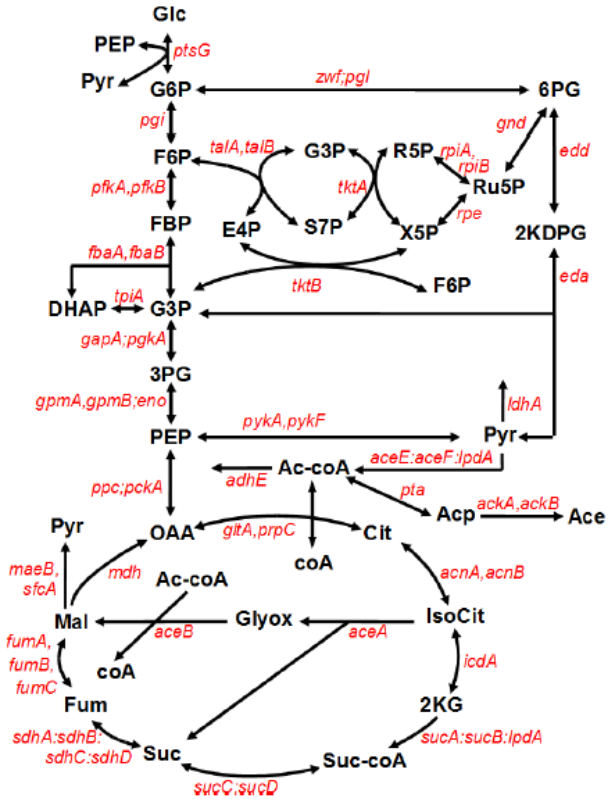
Identifiability analysis of *E. coli* metabolism

- Simplified model of central carbon metabolism in E. coli
- Model at steady state and lin-log kinetics for describing reactions

$$v = \mathrm{diag}(e) \cdot (a + B^x \cdot \ln x + B^u \cdot \ln u)$$
$$N \cdot v = 0$$

- Solution by minimal regression

  More complicated methods if data are incomplete



Berthoumieux *et al.* (2011), *Bioinformatics*, 27(13):i186–93

Identifiability analysis of *E. coli* metabolism

- Simplified model of central carbon metabolism in E. coli
- Most parameters are not identifiable as determined from profile likelihood!

Metabolites

Reactions	Glc	PEP	G6P	Pyr	F6P	FBP	DHAP	3PG	AcoA coA	6PG	Ru5P	R5P	S7P	2KG	Suc	Fum	Mal	ATP ADP	Cit	NADPH NADP	NADH NAD	FAD	Ace
	0.29	-0.89	0.79	1.87	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	-0.33	0	0.23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	-0.16	0	0	0.04	-0.28	0	0	0	0	0	0	0	0	0	0	0	0.09	0	0	0	0	0
	0	0	0	0	0	-0.3	0.06	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	-0.07	0.22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	-0.18	0	-0.05	0	0	0	0	0	0	0	0	0	0.32	0	0	-0.17	0	0
	0	0.26	0	0	0	0	0	-0.12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0.12	0	0.49	0	-0.19	0	0	0	0	0	0	0	0	0	0	0	0.16	0	0	0	0	0
	0	0	0	0.64	0	0	0	0	0.05	0	0	0	0	0	0	0	0	0	0	0	-0.21	0	0
	0	0	-0.22	0	0	0	0	0	0	-0.24	0	0	0	0	0	0	0	0	0	-0.01	0	0	0
	0	0	0	0	0	0	0	0	0	0.48	-0.09	0	0	0	0	0	0	0	0	-0.01	0	0	0
	0	0	0	0	0	0	0	0	0	0	0.46	0	-0.39	0	0	0	0	0	0	0	0	0	0
	0	0	-0.74	0	0	0	0	0	0	0	0.3	-0.16	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0.01	-0.1	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	-0.32	0	0	0	0	0	0	0	0.51	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	-0.58	0	0	0	0	0	0.35	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0.22	0	0	0	0	-0.001	0	0	0	0	0.49	0	-0.01	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0.99	0	0	0	0	0.55	0	0	0	0
	0	0	0	0	0	0	0	0	0.08	0	0	0	0	-0.09	0	0	0	0	0	-0.59	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	1.26	0.3	0	0	0	0	0	-0.48	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.08	-0.44	0	0	0	0	0	0.4	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.46	0.1	0	0	0	0	0	0
	0	0.15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.01	0	0.31	0	-0.12	0	0
	0	0.29	0	0	0	-0.13	0	0	0	0	0	0	0	0	0	0	0.1	-1.15	0.21	0	0	0	0
	0	0	0	-0.31	0	0	0	0	-0.21	0	0	0	0	0	0	0	0.38	0	0	0.36	-0.05	0	0
	0	0	0	0	0	0	0	0	-0.11	0	0	0	0	0	0.26	0	-0.18	0	0	0	0	0	0
	0	0.1	-0.09	0.04	-0.06	0	0	0.17	0.1	0	0	0.09	0	-0.01	0	0	0	-0.46	0	-0.003	0.01	0	0
	0	0	0	-0.03	0	0	0	0	0	-0.93	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	-0.06	0	0	0	0	-0.01	0	0	0	0	0	0	0	0	0.21	0	-0.25	-2.03	0	2.19
	0	0	0	2.67	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.11	0	0
	0	0	0	0	0	0	0	0	0.91	0	0	0	0	0	0	0	0	0	0	0	0	0	0

27/31 nonidentifiable reactions 73/100 nonidentifiable parameters

Berthoumieux *et al.* (2013), *J. Math. Biol.*, 67:1759-1832

Structural vs parametric identification

- **Parametric identification** assumes model structure given
- **Structural identification** provides joint estimate of model structure and parameters

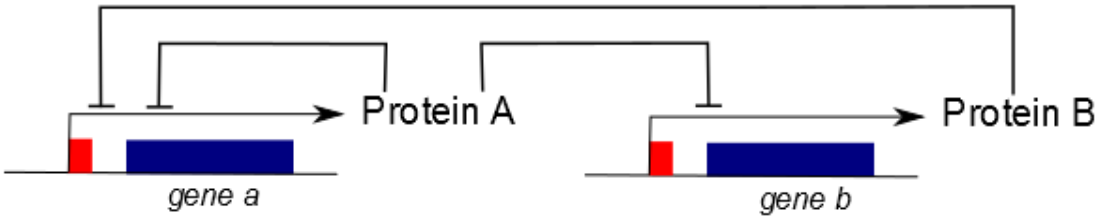
$$\mathbf{A}\frac{d}{dt}\mathbf{x}(t,\mathbf{p}) = \mathbf{f}(\mathbf{x}(t,\mathbf{p}),\mathbf{u}(t),\mathbf{p}), \quad t_0 \leq t \leq t_e$$
$$\mathbf{x}(t_0,\mathbf{p}) = \mathbf{x}_0$$

- Structural identification very difficult problem, based on exploration of *a-priori* defined model space

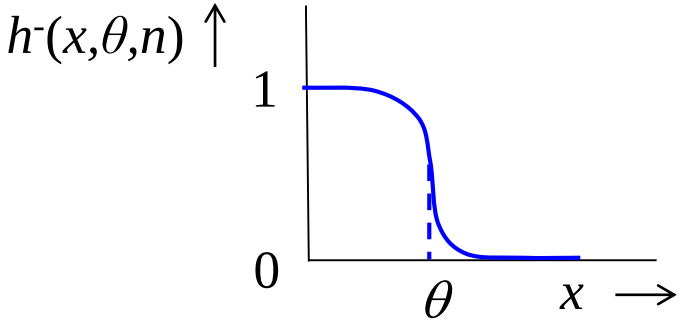
Structural vs parametric identification

- Method for **structural and parametric identification** of gene regulatory networks from time-series data
  - Exploits monotonicity properties of switch-like regulation functions *g* to invalidate interactions

$$\dot{x}_a = \kappa_a \, h^-(x_a, \theta_{a2}, n) \, h^-(x_b, \theta_b, n) - \gamma_a \, x_a$$
$$\dot{x}_b = \kappa_b \, h^-(x_a, \theta_{a1}, n) - \gamma_b \, x_b$$



Porreca *et al.* (2010), *Bioinformatics*, 26(9):1239-45



x : protein concentration
θ : threshold concentration
κ, γ : rate constants
n : steepness parameter

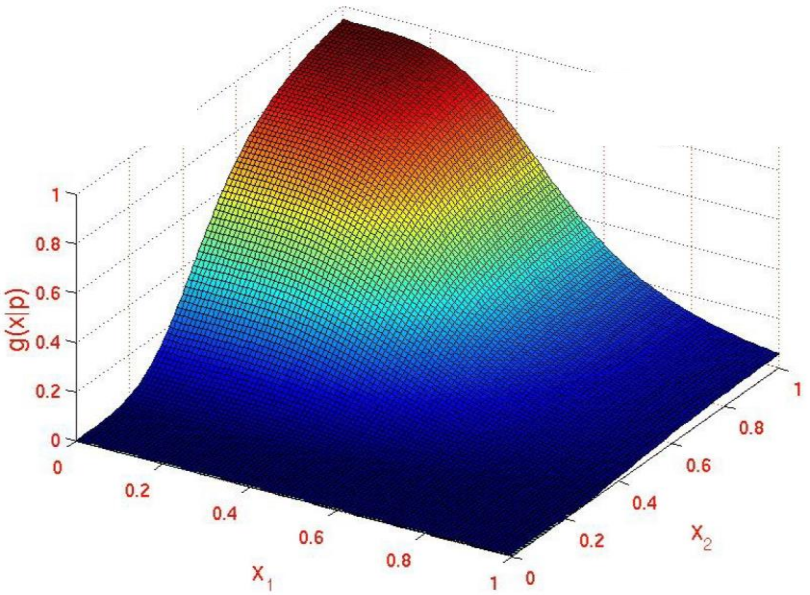
Structural vs parametric identification

- Method for **structural and parametric identification** of gene regulatory networks from time-series data
  - Exploits monotonicity properties of switch-like regulation functions *g* to invalidate interactions

Observations $x_1 \uparrow$, $x_2 \downarrow$ and $g(x_1, x_2) \uparrow$ rule out pattern x_1 activator and x_2 inhibitor

$$\dot{x}_3 = g(x_1, x_2, p) - \gamma_3 x_3$$

$$g(x_1, x_2, p) = \kappa_3 h^+(x_1, \theta_1, n)\, h^-(x_2, \theta_2, n)$$

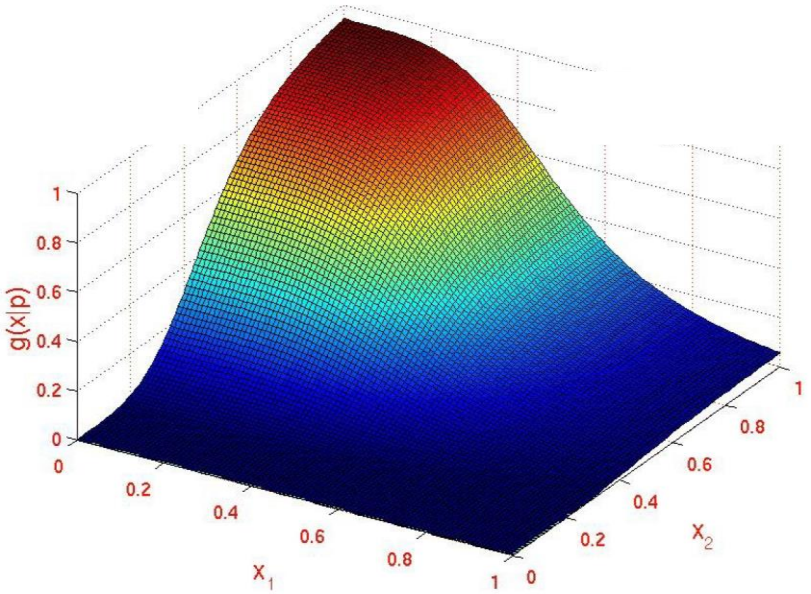


Porreca *et al.* (2010), *Bioinformatics*, 26(9):1239-45

Structural vs parametric identification

- Method for **structural and parametric identification** of gene regulatory networks from time-series data
  - Exploits monotonicity properties of switch-like regulation functions *g* to invalidate interactions
  - Looks for simplest interaction structures consistent with data

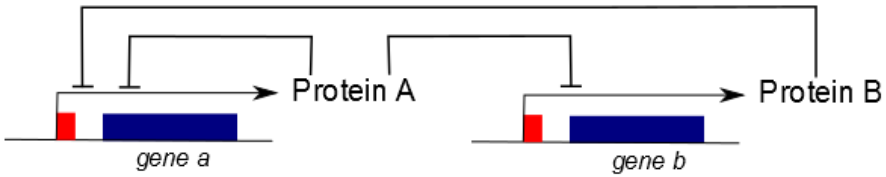
Subpatterns of inconsistent patterns are also inconsistent
Superpatterns of consistent patterns are also consistent

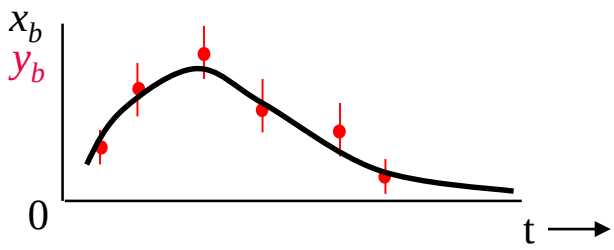


Porreca *et al.* (2010), *Bioinformatics*, 26(9):1239-45

Structural vs parametric identification

- Method for **structural and parametric identification** of gene regulatory networks from time-series data
  - Exploits monotonicity properties of switch-like regulation functions *g* to invalidate interactions
  - Looks for simplest interaction structures consistent with data
  - Estimate parameters for selected interaction structure

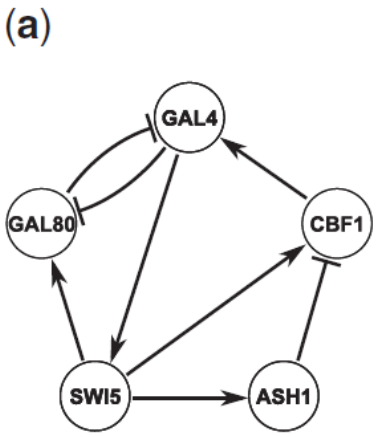




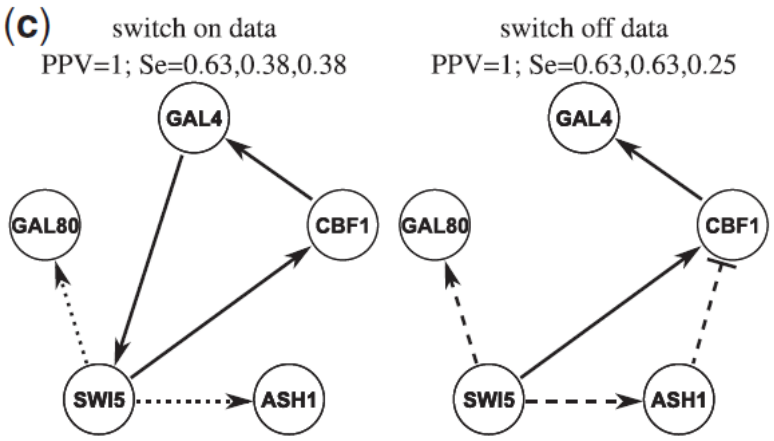
Porreca *et al.* (2010), *Bioinformatics*, 26(9):1239-45

Structural vs parametric identification

- Method for **structural and parametric identification** of gene regulatory networks from time-series data
  - Exploits monotonicity properties of switch-like regulation functions *g* to invalidate interactions
  - Looks for simplest interaction structures consistent with data
  - Estimate parameters for selected interaction structure
  - Good results on benchmark data set in yeast



True network



Algorithm 1

Porreca *et al.* (2010), *Bioinformatics*, 26(9):1239-45

Structural vs parametric identification

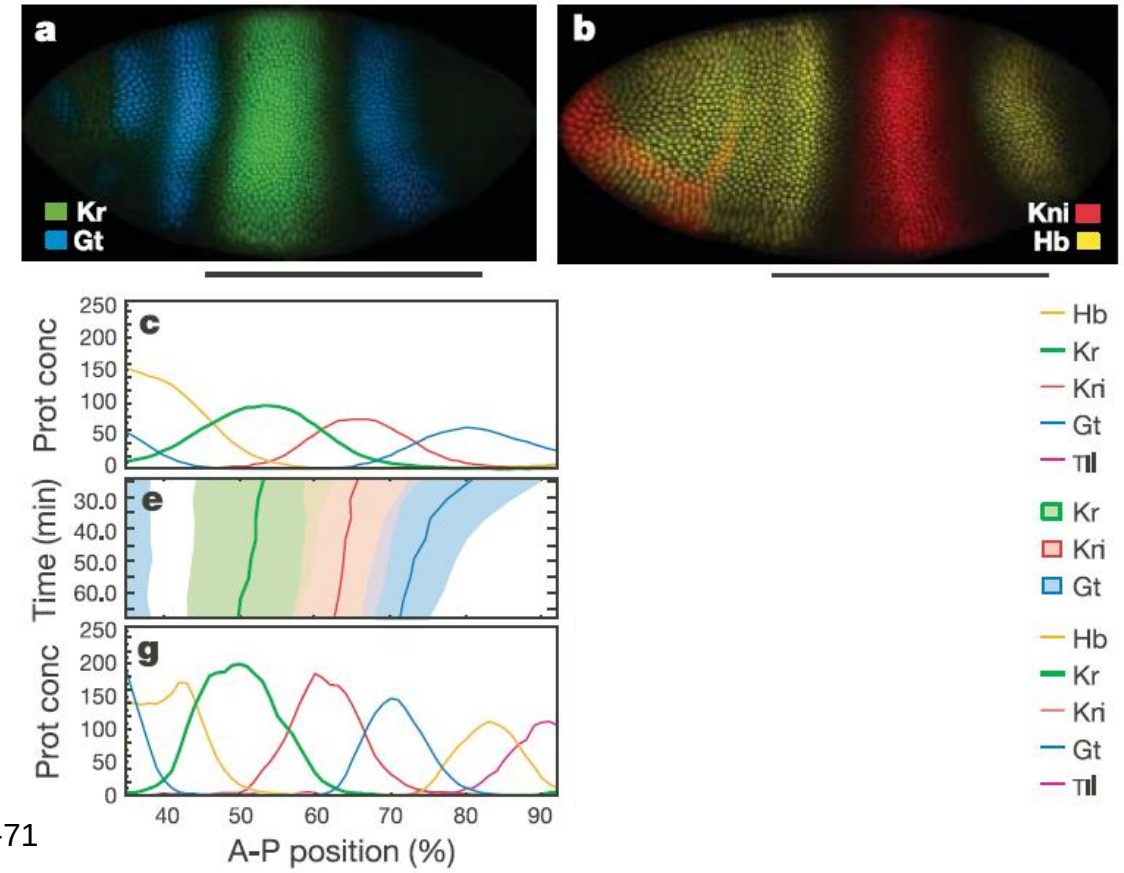
- **Parametric identification** assumes model structure given
- **Structural identification** provides joint estimate of model structure and parameters

$$\mathbf{A}\frac{d}{dt}\mathbf{x}(t,\mathbf{p}) = \mathbf{f}(\mathbf{x}(t,\mathbf{p}),\mathbf{u}(t),\mathbf{p}), \quad t_0 \leq t \leq t_e$$
$$\mathbf{x}(t_0,\mathbf{p}) = \mathbf{x}_0$$

- Structural identification very difficult problem, based on exploration of *a-priori* defined model space
- In certain cases, structural identification reduces to parametric identification

Parameter estimation from *Drosophila* data

- Measurement of protein concentrations of gap genes during development of *Drosophila* embryon



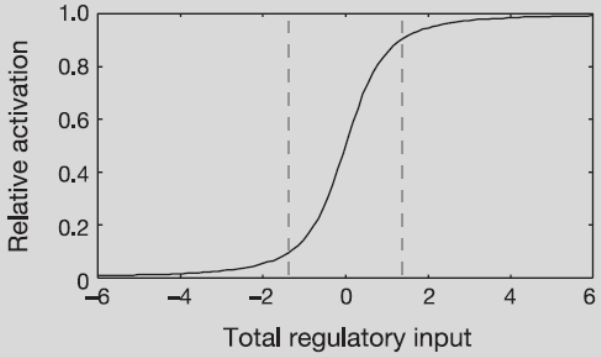
Jaeger *et al.* (2004), *Nature*, 430(6997):368-71

Parameter estimation from *Drosophila* data

- Neural-network-like model of connections between gap genes
  - Model with 58 nuclei and 7 variables (proteins) per nucleus
  - Free diffusion of proteins because at early stages of development embryon is syncytium (multinucleate cell)
  - Sigmodial response functions
  - Connectivity pattern encoded in parameter matrix *T*, so parametric **and** structural identification

$$\frac{dv_i^a}{dt} = R_a g(u^a) + D^a\left[(v_{i-1}^a - v_i^a) + (v_{i+1}^a - v_i^a)\right] - \lambda_a v_i^a$$

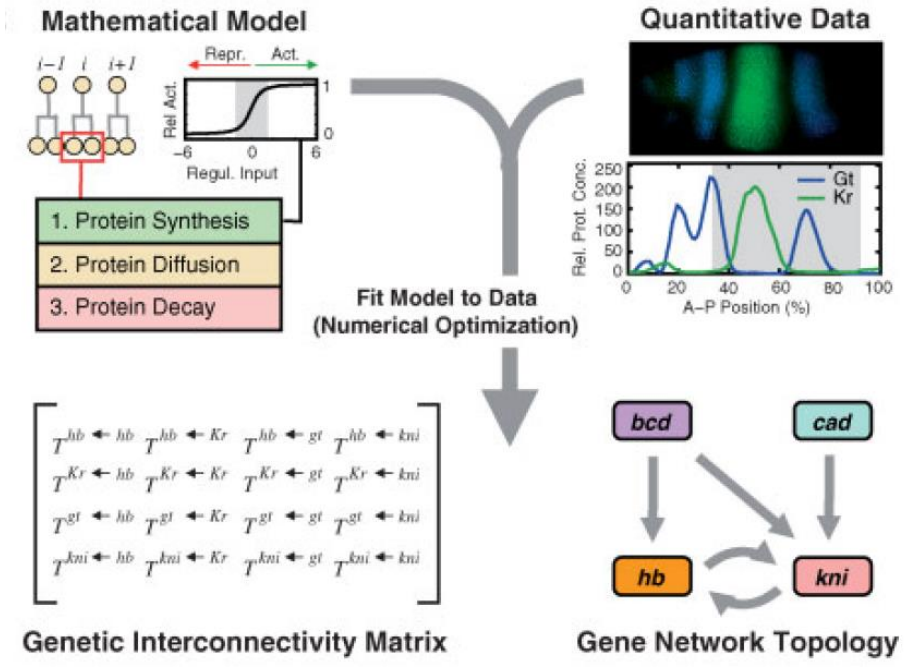
$$u^a = \sum_b T^{ab} v_i^b + m^a v_i^{\mathrm{Bcd}} + h^a$$



Jaeger *et al.* (2004), *Nature*, 430(6997):368-71

Parameter estimation from *Drosophila* data

- Neural-network-like model of connections between gap genes
- Brute-force parameter estimation by fitting model to data
  - Parallelized simulated annealing

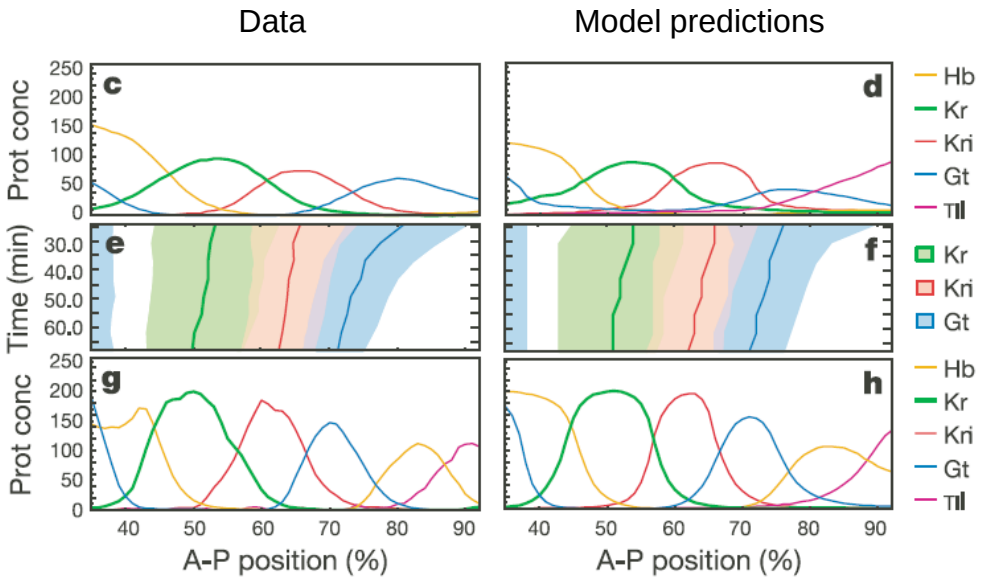


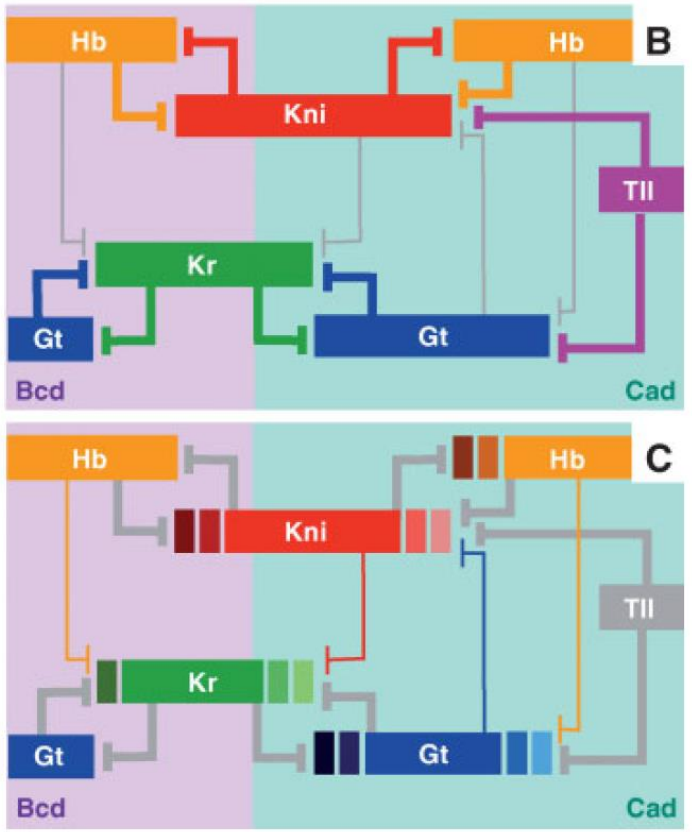
Jaeger and Reinitz (2006), *BioEssays*, 28(11):1102-11

Shifts in gap gene domains

- What is function of **cross-inhibition between gap genes**?

  Model predicts that they are important for shift in gap gene domains after their initial establishment





Conclusions

- Identification of models requires estimation of parameters
  - Optimization problem: minimization of objective function
  - Large variety of methods available: global vs local, deterministic vs stochastic, …
- Identifiability issues: structural and practical identifiability
- Structural vs parametric identification
- Other issues: optimal experimental design, ensemble models
- Large-scale parameter estimation problems in systems biology are very difficult to solve
  - Clever tricks
  - Model-dependent heuristics
  - Model reduction, …

Internships in IBIS

- Challenging problems for biologists, physicists, computer scientists, mathematicians, …
- … in a multidisciplinary working environment
- Contact: Hidde.de-Jong@inria.fr and ibis.inrialpes.fr

Courtesy Guillaume Baptist (2008)

Merci !

team.inria.fr/ibis