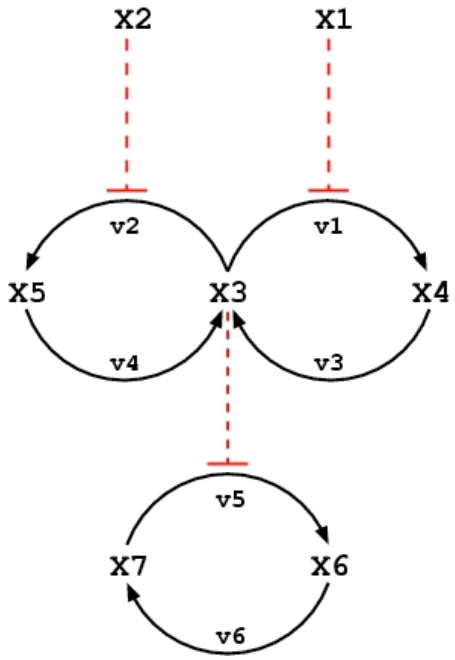
# Structural identification of dynamical systems - overview and reflection

LCCC Workshop 2010, Lund

Dag Wedelin and Peter Gennemark
Chalmers University of Technology and University of Göteborg

Work supported by the Göteborg Mathematical Modelling Centre

# Classical modelling of biomolecular systems




Qualitative static models

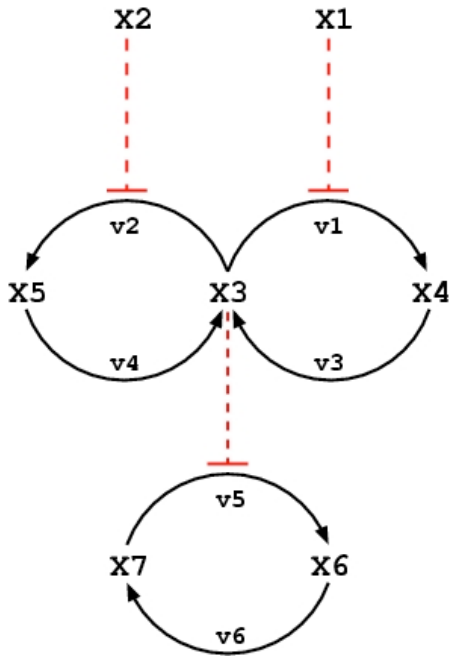
$$
\begin{aligned}
S_3'(t) &= -v_1 - v_2 + v_3 + v_4 \\
S_4'(t) &= v_1 - v_3 \\
S_5'(t) &= v_2 - v_4 \\
S_6'(t) &= -S_7'(t) = v_5 - v_6
\end{aligned}
$$

$$
\begin{aligned}
v_1 &= \frac{S_3(t)V_{\max 1}}{(S_3(t)+K_{D1})(1+I_1(t)/K_{I1})} \\
v_2 &= \frac{S_3(t)V_{\max 2}}{(S_3(t)+K_{D2})(1+I_2(t)/K_{I2})} \\
v_3 &= \frac{S_4(t)V_{\max 3}}{S_4(t)+K_{D3}} \\
v_4 &= \frac{S_5(t)V_{\max 4}}{S_5(t)+K_{D4}} \\
v_5 &= \frac{S_7(t)V_{\max 5}}{(S_7(t)+K_{D5})(1+S_3(t)/K_{I3})} \\
v_6 &= \frac{S_6(t)V_{\max 6}}{S_6(t)+K_{D6}}
\end{aligned}
$$

Manually designed ODE models

(metabolic pathway system, Arkin & Ross 1995)

# Possible goals of structural identification




Sketch the network structure

Find models with explanatory value!

$$S_3'(t) = -v_1 - v_2 + v_3 + v_4$$
$$S_4'(t) = v_1 - v_3$$
$$S_5'(t) = v_2 - v_4$$
$$S_6'(t) = -S_7'(t) = v_5 - v_6$$

$$v_1 = \frac{S_3(t)V_{\max 1}}{(S_3(t) + K_{D1})(1 + I_1(t)/K_{I1})}$$
$$v_2 = \frac{S_3(t)V_{\max 2}}{(S_3(t) + K_{D2})(1 + I_2(t)/K_{I2})}$$
$$v_3 = \frac{S_4(t)V_{\max 3}}{S_4(t) + K_{D3}}$$
$$v_4 = \frac{S_5(t)V_{\max 4}}{S_5(t) + K_{D4}}$$
$$v_5 = \frac{S_7(t)V_{\max 5}}{(S_7(t) + K_{D5})(1 + S_3(t)/K_{I3})}$$
$$v_6 = \frac{S_6(t)V_{\max 6}}{S_6(t) + K_{D6}}$$

Reconstruct complete model

- interaction structure
- structure of equations
- parameters

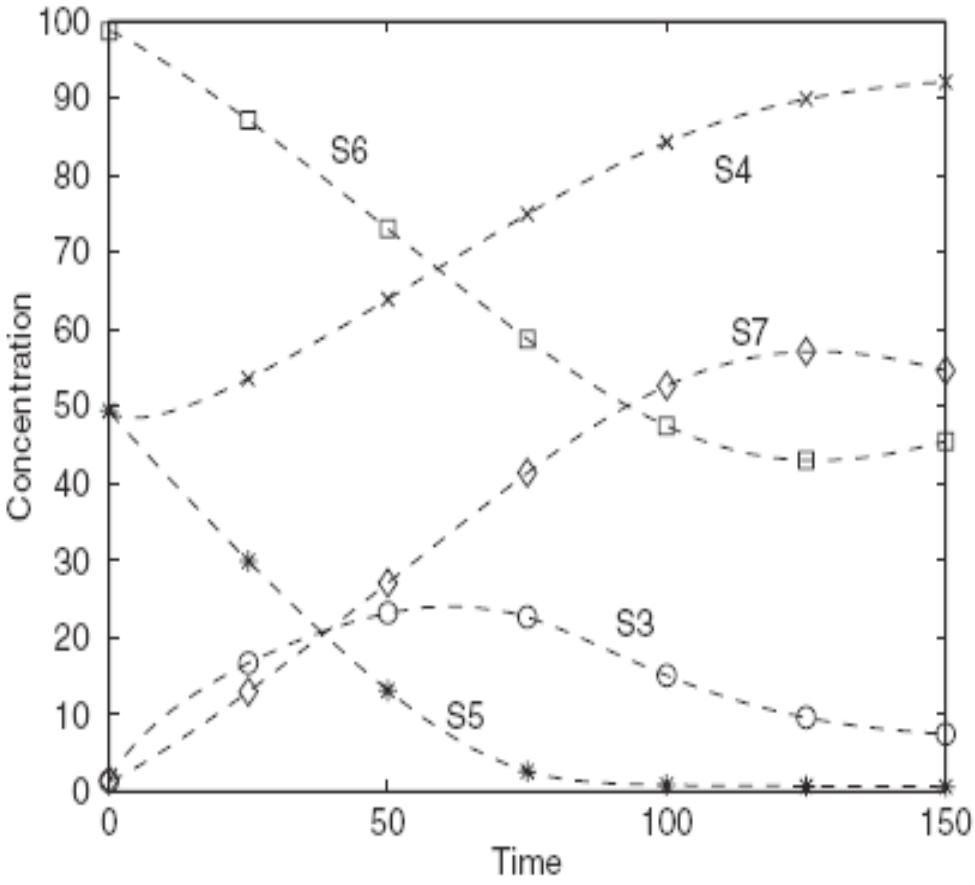
# Input to structural identification problems considered in the biological domain

Time series data for all variables (steady state data, ...)

Sparse, irregular, noisy data

There can be several experiments

Knockouts



# Common model types for biomolecular systems

$$X_3' = 0.2\, x_1\, x_2' - 0.4\, \frac{x_1\, x_3}{x_3 + 2.1}$$

chemical rate reactions

$$X_3' = 0.36\, x_4^{0.21} + 0.23\, x_2^{0.32} - 0.73\, x_2^{0.85}\, x_7^{0.4}$$

GMA

$$X_6' = 0.8\, x_4^{0.31}\, x_5^{0.2} - 0.12\, x_2^{0.11}\, x_5^{-0.24}$$

S-system - special case of GMA

non-linear parameter estimation

# Current algorithmic research in this area

METHODS

Usually S-system (fixed form)

Often lack of clear problem statement

Typically so called evolutionary computation

Parameter estimation of entire problem in inner loop

RESULTS

Large amounts of data required.

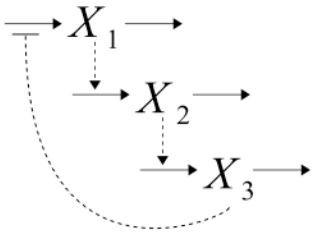
High to extreme computation times, considered as supercomputing topic.

Common published result: it was possible to reconstruct this single system

A relatively young topic in this area of application.

# How define a structural identification problem?

## A. THE SYSTEM

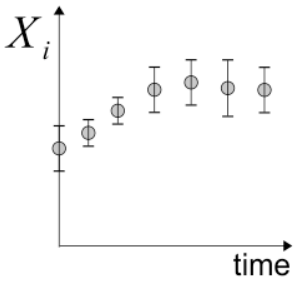



$$X_1'(t) = \frac{0.81}{X_3(t) + 0.9} - 1.0 * X_1(t)$$

$$X_2'(t) = 1.0 * X_1(t) - 0.6 * X_2(t)$$

$$X_3'(t) = 0.6 * X_2(t) - 0.8 * X_3(t)$$

## B. THE PROBLEM




Data

$$v_1 = k_1 X_i \qquad i \in [1,3], \quad k_1 \in [0,50]$$

$$v_2 = \frac{k_1 X_i}{X_i(t) + k_2} \qquad i \in [1,3], \quad k_{1,2} \in [0,50]$$

$$v_3 = \frac{k_1 k_2}{X_i(t) + k_2} \qquad i \in [1,3], \quad k_{1,2} \in [0,50]$$

Model space and constraints

$$Error = -L + \lambda K$$

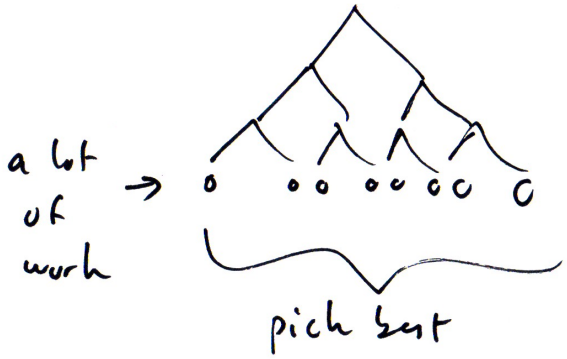
Error function

$$X_1'(t) = X_2'(t) = X_3'(t) = 0$$

Initial model

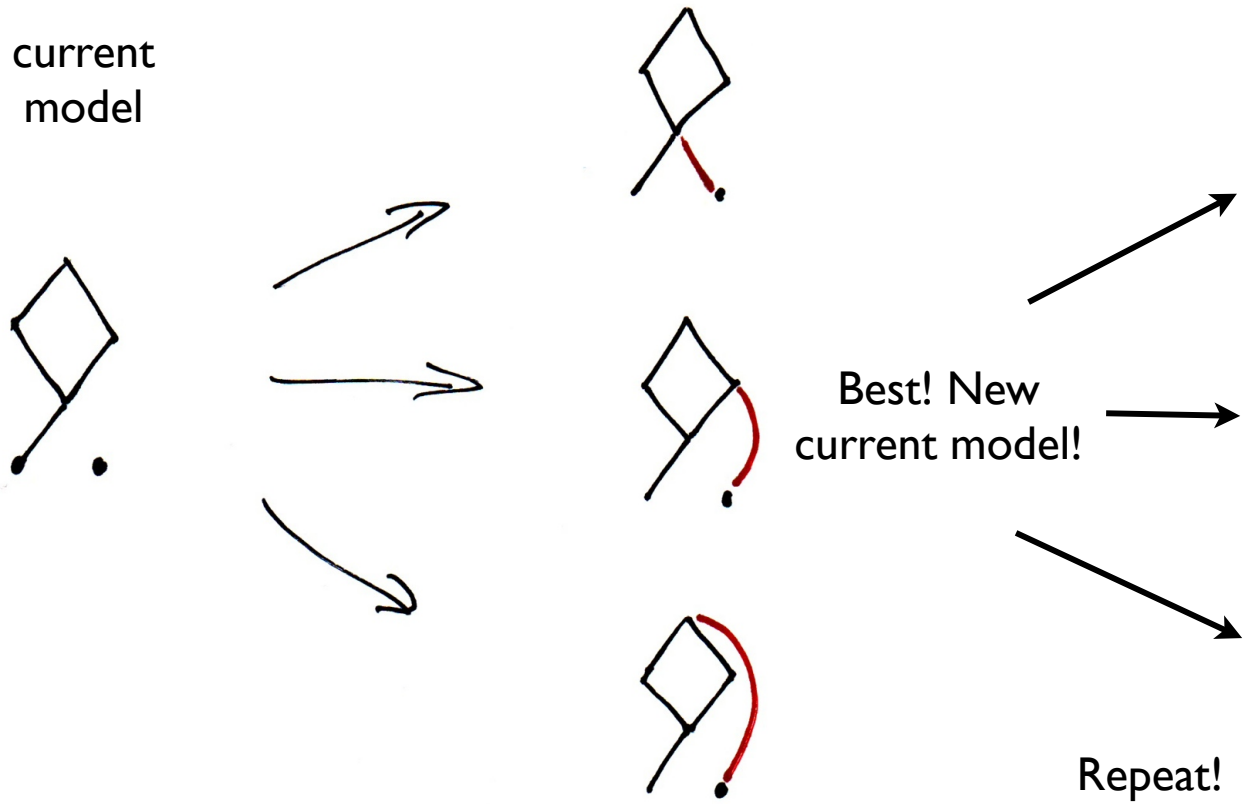
Problem fully defined as an optimization problem!

# The slowest algorithm



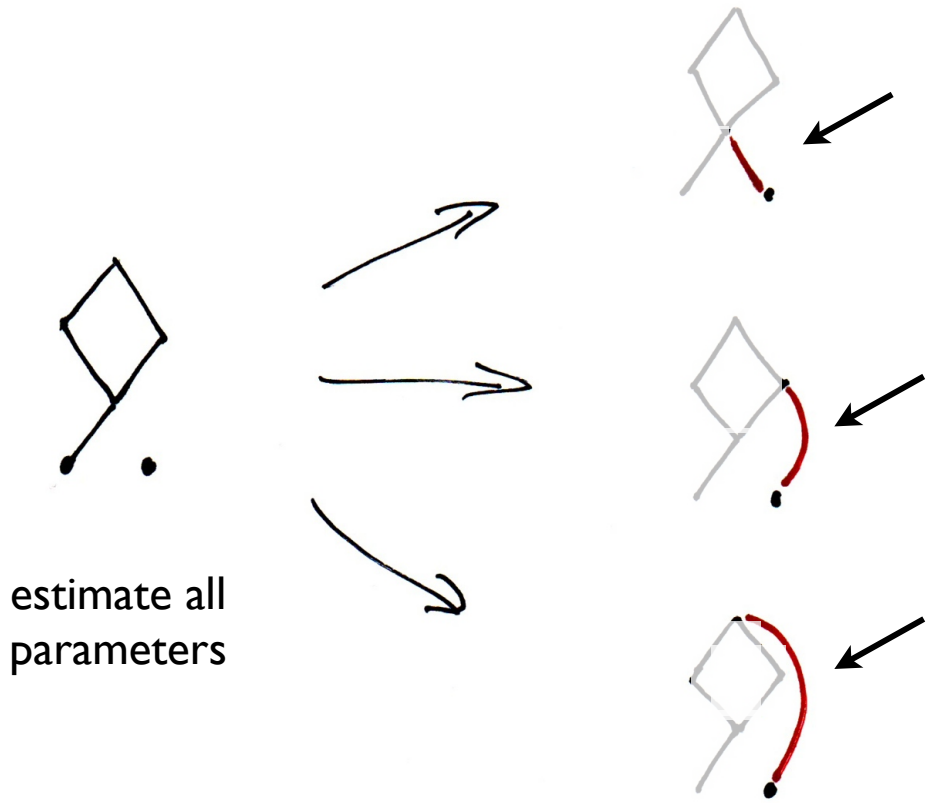

1. Enumerate all possible structures
2. For every structure, estimate its parameters
3. Pick the solution with the smallest error

# Avoid complete enumeration with local search




local operations in the network structure!

# Decompose parameter estimation

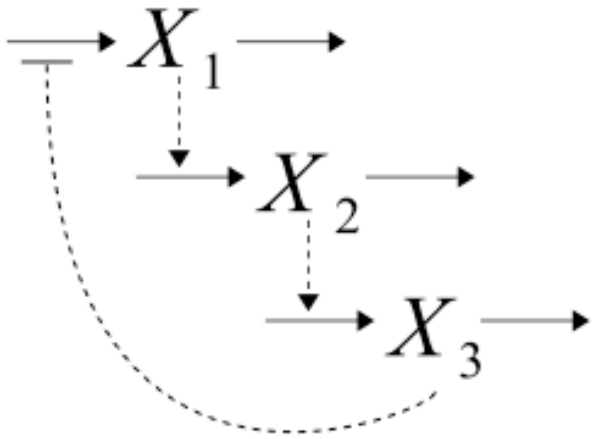


estimate all parameters

estimate only new parameters!

local operations in the network structure!

# A simple ODE model

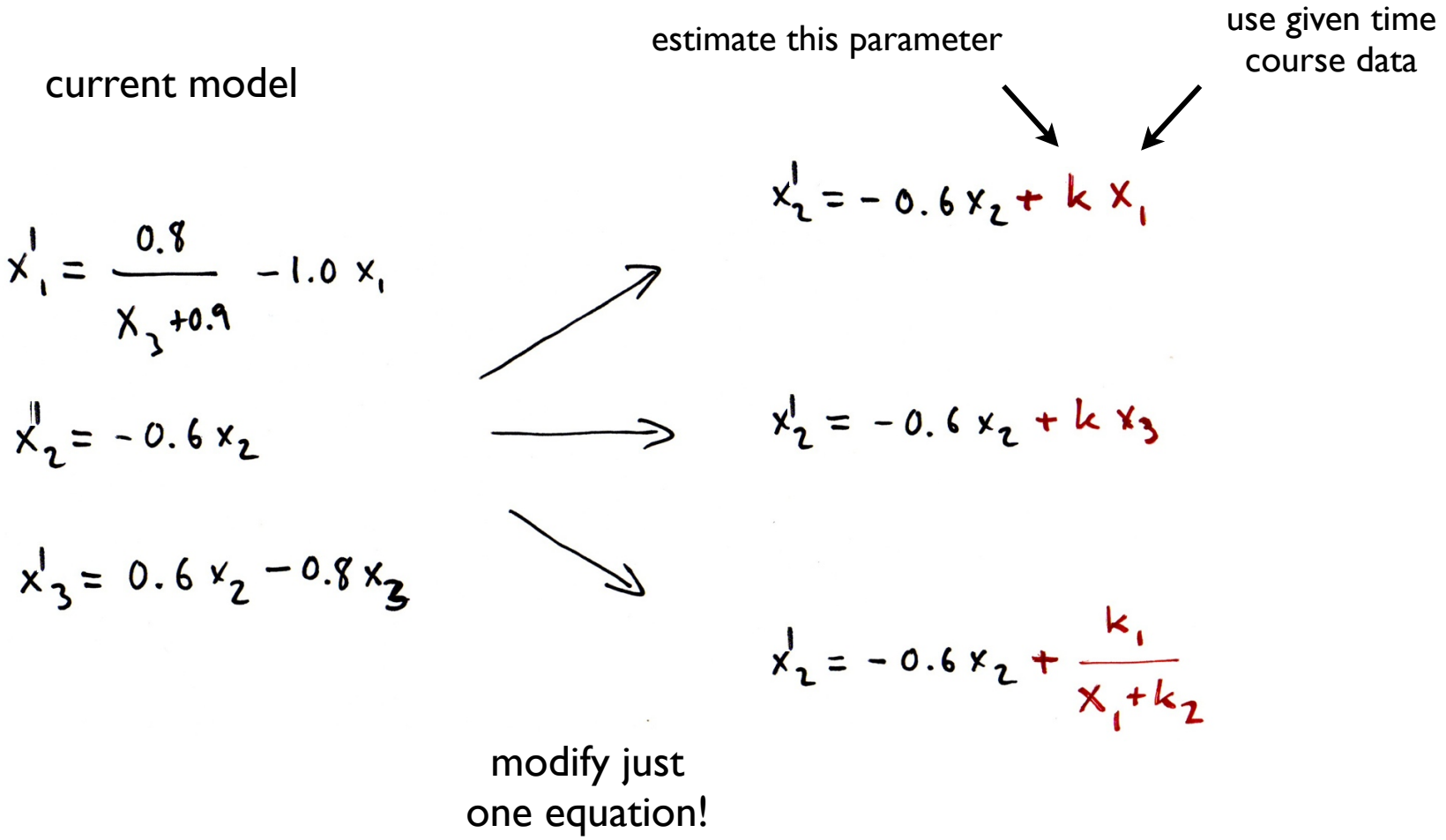



$$X_1'(t) = \frac{0.81}{X_3(t) + 0.9} - 1.0 * X_1(t)$$

$$X_2'(t) = 1.0 * X_1(t) - 0.6 * X_2(t)$$

$$X_3'(t) = 0.6 * X_2(t) - 0.8 * X_3(t)$$

# ODE model selection algorithm



# Single equation parameter estimation

estimate from given time course data

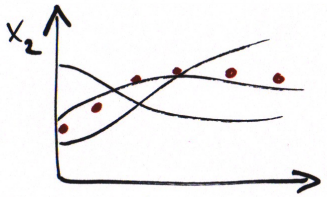
use given time course data

$$x_2' = k_1 x_2 + k_2 x_1$$

rough

use given time course data

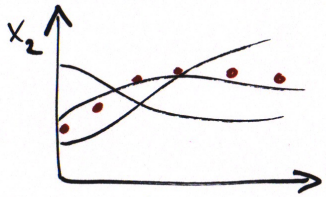
$$x_2' = k_1 x_2 + k_2 x_1$$




more exact

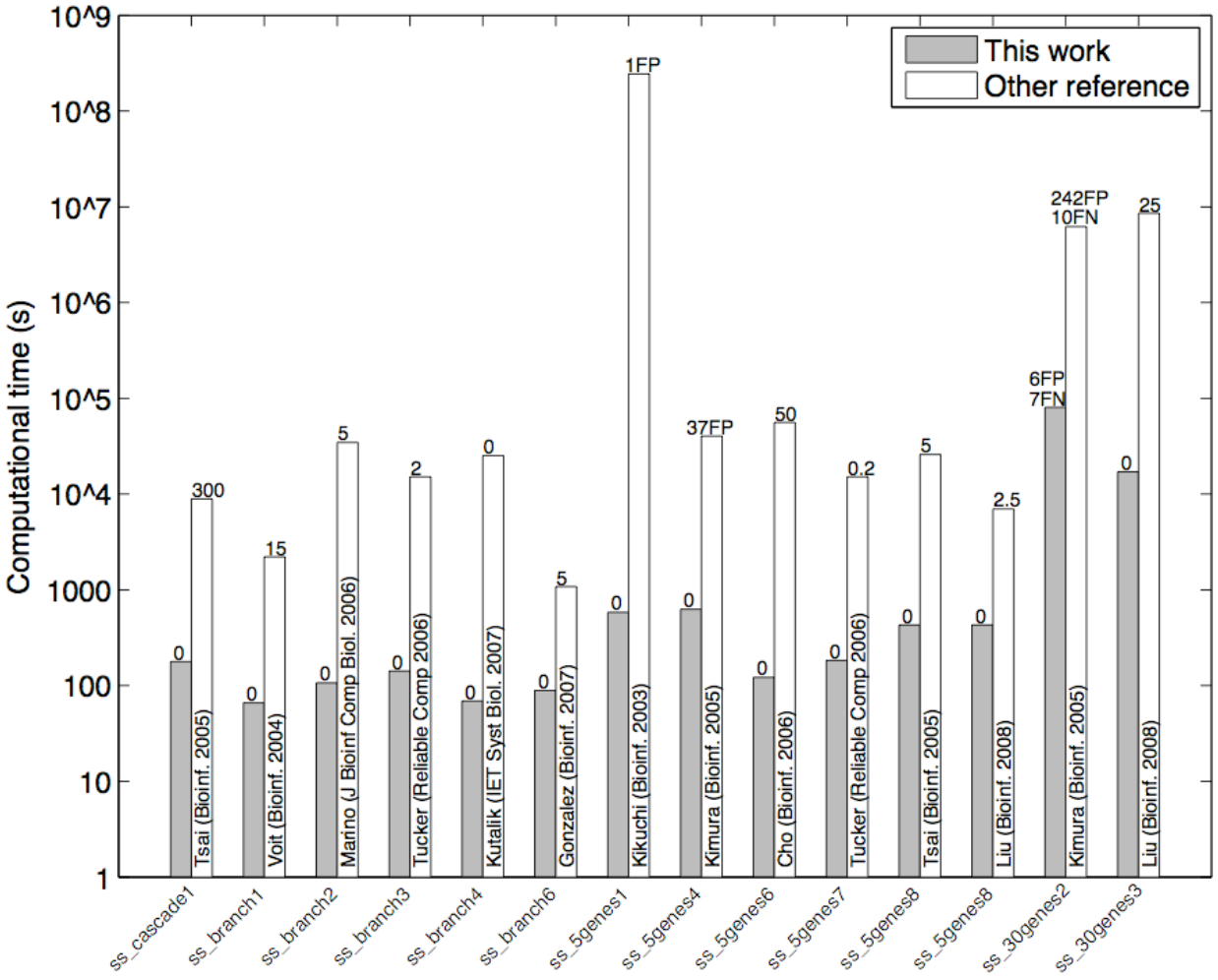
use simulated data

$$x_2' = k_1 x_2 + k_2 x_1$$




most exact

# Speed and accuracy



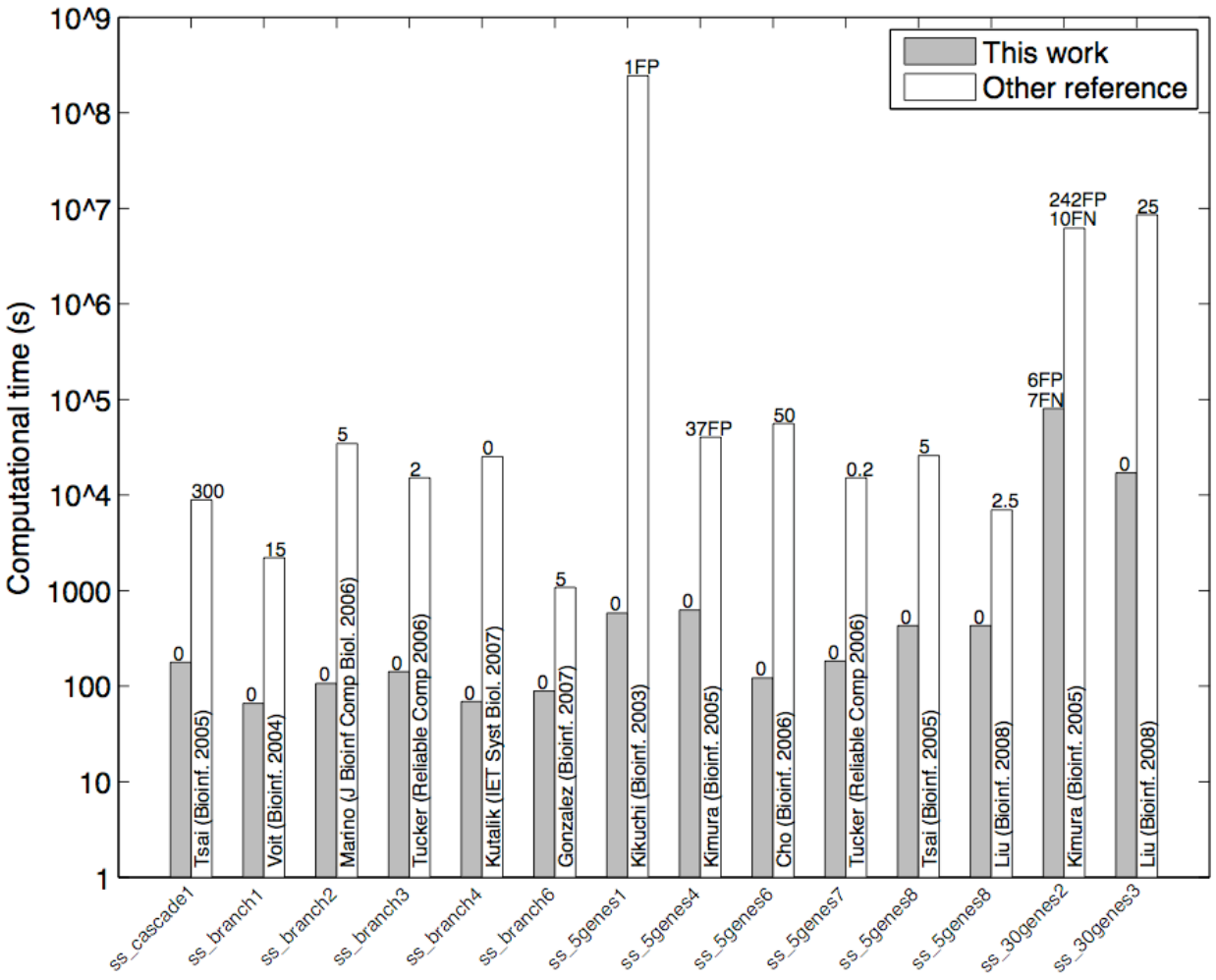

(see odeidentification.org for benchmark problems and results)

We do not think that speed and accuracy is the bottleneck in practice.

We require reasonable amounts of data.

MANY variables (will be) possible.

# Speed and accuracy




(see odeidentification.org for benchmark problems and results)

We do not think that speed and accuracy is the bottleneck in practice.

We require reasonable amounts of data.

MANY variables (will be) possible.

Still gap to biological practice...

# Different kinds of structural identification

Structural function identification (or regression):

$$f(x) = -3.4x + 0.11x^4$$

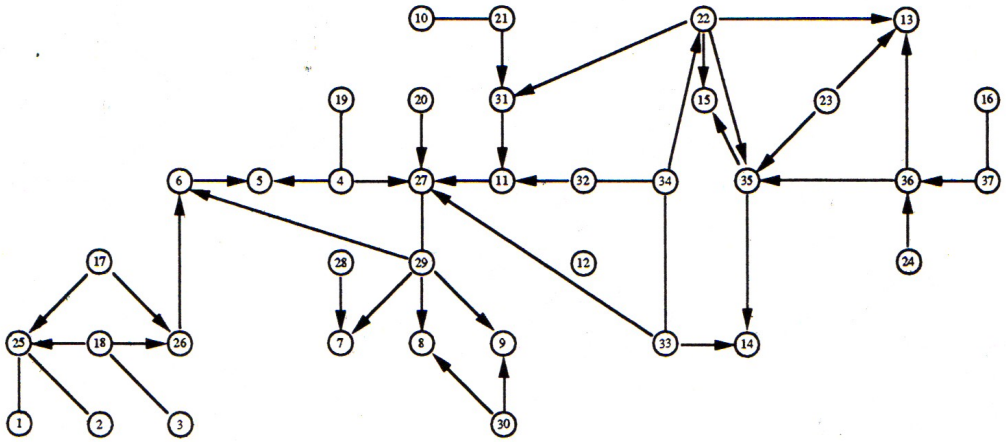
Which terms should be selected?

Structural identification of:

- dynamical systems - function identification if perfect data
- probabilistic graphical models - example of density estimation, structural function identification if perfect (infinite) data
- equations - find non-trivial invariant function

Different error functions for different problem types.

# Local algorithmic approach for probabilistic graphical models

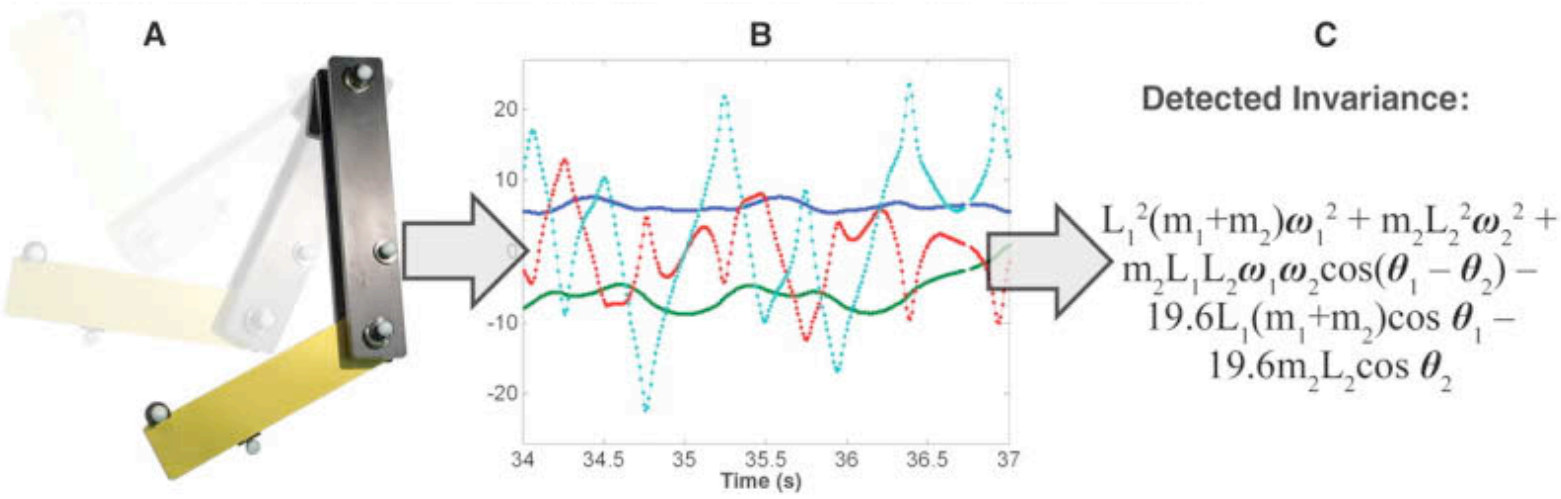



Local structure search and parameter estimation approach is useful also for this kind of networks.

All details different...

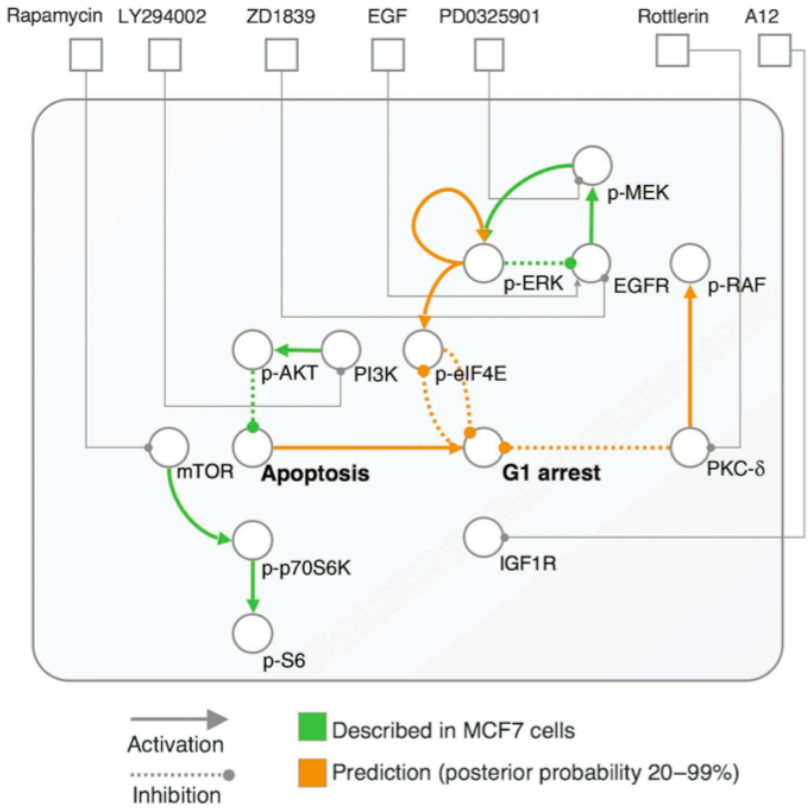
Factorize the joint probability distribution!

# Distilling Free-Form Natural Laws from Experimental Data (Schmidt, Lipson 2009)




”symbolic” regression

# Models from experiments: combinatorial drug perturbations of cancer cells (Nelander et al 2008)




$$
\begin{cases}
\frac{d\text{IGF1R}}{dt} = \phi(-\text{A12}) - \text{IGF1R} \\
\frac{d\text{PI3K}}{dt} = 1.14\phi(+0.05\text{p-eIF4E} - \text{LY294002}) - 0.84\text{PI3K} \\
\frac{d\text{mTOR}}{dt} = 1.04\phi(-\text{Rapamycin}) - 0.96\text{mTOR} \\
\frac{d\text{PKCdelta}}{dt} = 1.05\phi(-\text{Rottlerin}) - 0.95\text{PKCdelta} \\
\frac{d\text{MEK}}{dt} = 0.63\phi(-0.21\text{MEK} + 0.76\text{EGFR} - \text{PD901}) - 1.27\text{MEK} \\
\frac{d\text{EGFR}}{dt} = 1.25\phi(-0.34\text{p-ERK1/2} + 0.37\text{G1arrest} - \text{EGF}) - 0.66\text{EGFR} \\
\frac{d\text{p-ERK1/2}}{dt} = 1.13\phi(+0.41\text{MEK} + 0.32\text{p-ERK1/2} - 0.13\text{G1arrest}) - 0.86\text{p-ERK1/2} \\
\frac{d\text{p-AKT}}{dt} = \phi(+0.55\text{PI3K})\text{-p-AKT} \\
\frac{d\text{p-P70S6K}}{dt} = 1.12\phi(+0.28\text{mTOR} - 0.41\text{Apoptosis}) - 0.87\text{p-P70S6K} \\
\frac{d\text{p-eIF4E}}{dt} = 1.07\phi(+0.21\text{p-ERK1/2} - 0.33\text{G1arrest}) - 0.92\text{p-eIF4E} \\
\frac{d\text{p-c-Raf}}{dt} = 1.08\phi(+0.29\text{PKCdelta} + 0.3\text{MEK} - 0.07\text{EGFR}) - 0.91\text{p-c-Raf} \\
\frac{d\text{p-S6}}{dt} = 1.13\phi(+0.52\text{p-P70S6K}) - 0.85\text{p-S6} \\
\frac{d\text{G1arrest}}{dt} = 1.11\phi(-0.05\text{IGF1R} - 0.14\text{PKCdelta} \\
\qquad -0.24\text{p-eIF4E} + 0.4\text{Apoptosis}) - 0.87\text{G1arrest} \\
\frac{d\text{Apoptosis}}{dt} = 1.09\phi(-0.06\text{mTOR} - 0.42\text{p-AKT}) - 0.91\text{Apoptosis} \\
\text{where } \phi(x) = \tanh(2x)
\end{cases}
$$

Confirmation of known interactions. ”Prediction” of new interactions.

# To discuss...

## METHODS

How specify structural identification problems?

General models and problem types?

What is possible in theory, and how will it affect practice?

Same inductive principles for structural identification?

What is the explanatory value?

Terminology?

## APPLICATION IN BIOLOGY

Still gap to biological practice, because of the data problem...

## GENERAL

Is automatic structural identification realistic and useful?

We can simulate the entire discovery process!

Representative test problems from different domains?

...

END OF TALK