Identifiability Analysis and Experimental Design for Dynamical Models in Systems Biology

Jens Timmer

Center for Systems Biology
Center for Data Analysis and Modeling
Center for Applied Biosciences
Bernstein Center for Computational Neuroscience
Freiburg Institute for Advanced Studies
Department of Mathematics and Physics
University of Freiburg

http://www.fdm.uni-freiburg.de/∼jeti/

Outline

- Systems Biology
- (Non-)Identifiability
- A New Method

Enlarging Math, Physics, Engineering

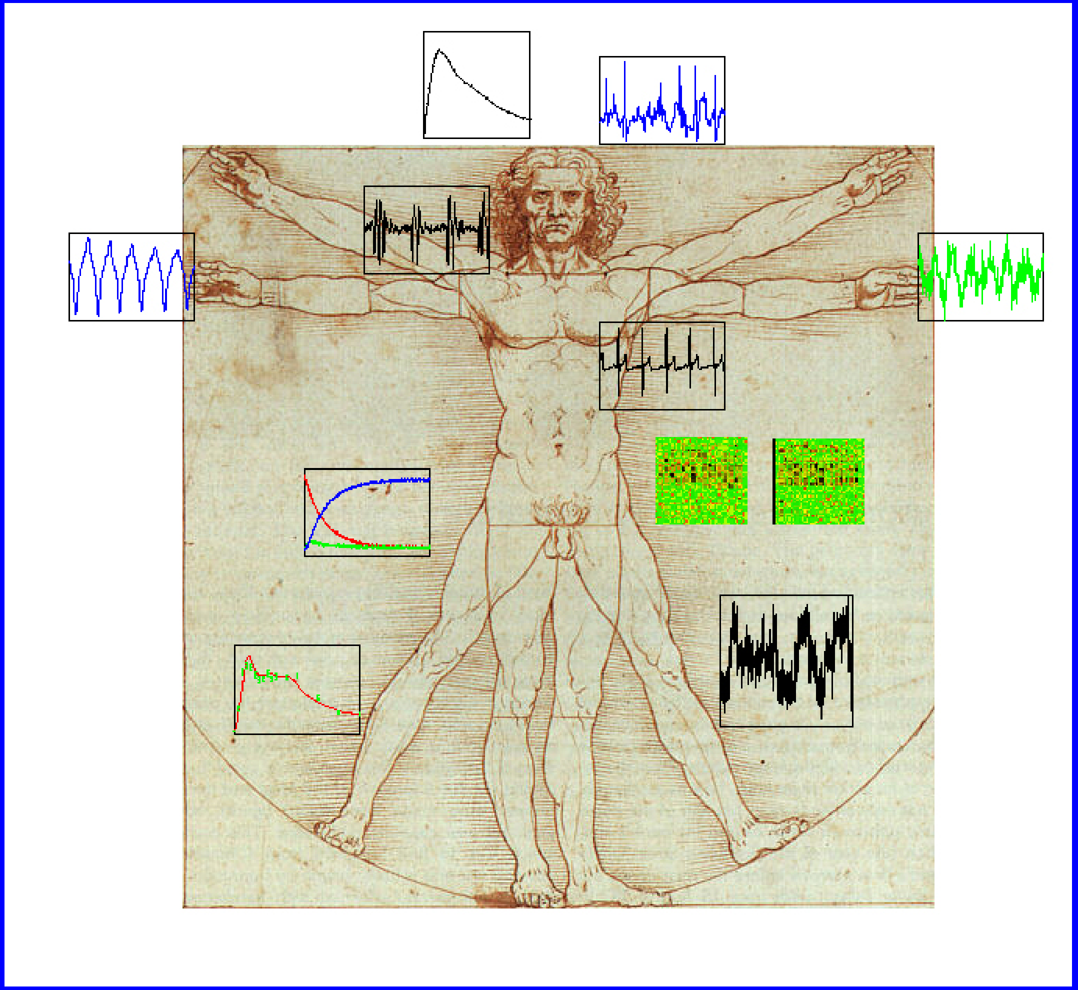
- **Since Newton:**

  Mathematatization of inanimate nature

- **21st century:**

  Additionally: Mathematization of animate nature

Man : A Dynamical System



Diseases caused or expressed by malfunction of dynamical processes

Two Directions in Systems Biology

- **Putting all the omics together**

  So far: large scale, qualitative, static

- **Understanding biomedical networks by data-based mathematical modelling of their dynamical behavior**

  So far: small scale, quantitative, dynamic

Both approaches will converge to: large scale, quantitative, dynamic

Common ground: Investigating networks

Our Direction in Systems Biology

Understanding biomedical systems by data-based

mathematical modelling of their dynamical behavior

From components and structure to behavior of networks

Systems Biology is based on but more than ...

- ... Mathematical Biology: Data-based
- ... Bioinformatics: Dynamics
- ... o.p./g. – o.p.: System
- ... another omics: Mathematics

Why Mathematical Modelling in BioMed?

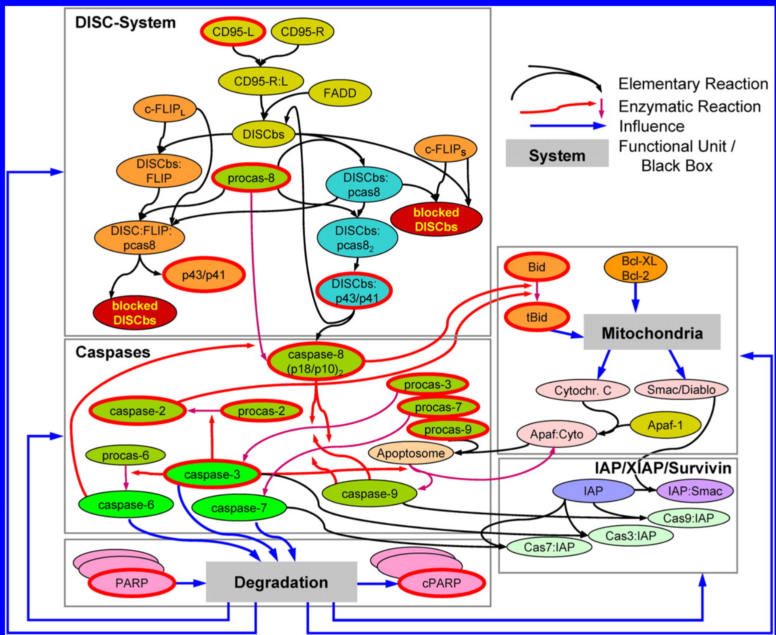
- Make assumptions explicit
- Understand essential properties, failing models
- Condense information, handle complexity
- Understand role of dynamical processes, e.g. feed-back
- Impossible experiments become possible
- Prediction and control
- Understand what is known
- Discover general principles
- "You don't understand it until you can model it"

Why Modelling in Cell Biology?

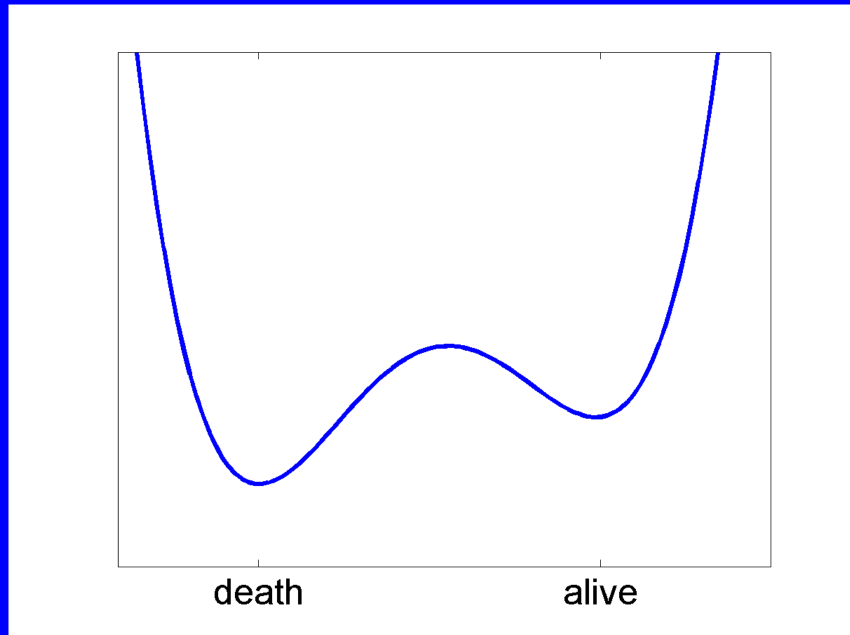
- Basic Research
  - Genomes are sequenced, but ...
  - ... function determined by regulation
  - Regulation = Interaction & Dynamics
  - Function: Property of dynamic network
  - "Systems Biology"
- Application
  - Drug development takes 10 years and 1 bn $/€
  - Reduce effort by understanding systems

Examples of Networks I: Apoptosis

Pathway cartoon



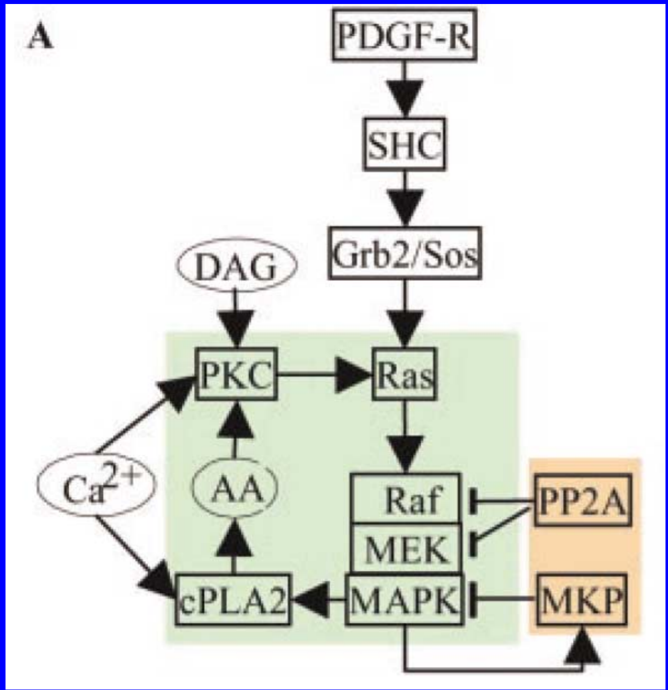
System's behavior



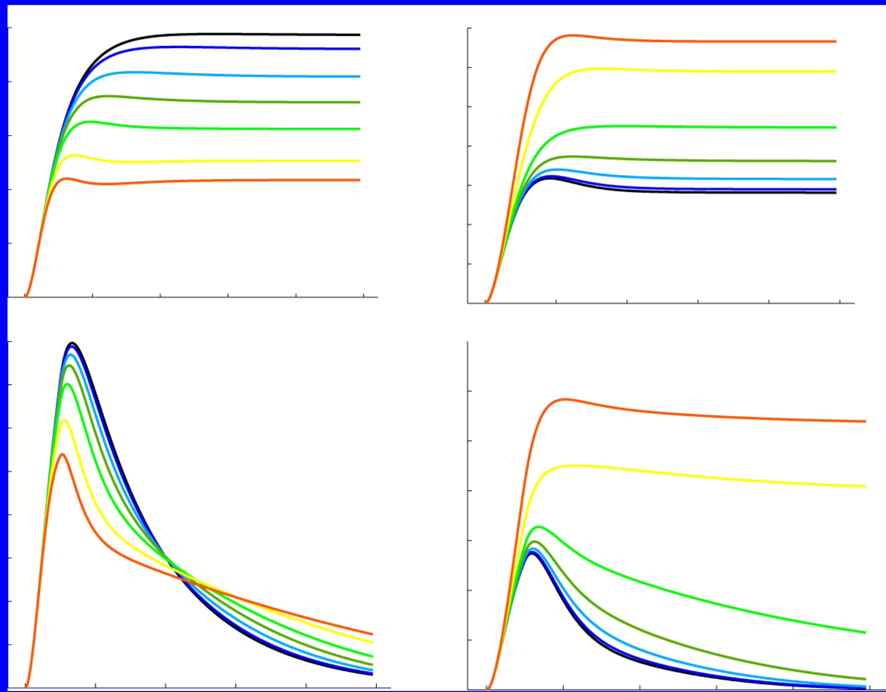
Threshold behavior, one-way bistable

Examples of Networks II: MAP Kinase

Pathway cartoon



System's behavior



Time scales/parameters important

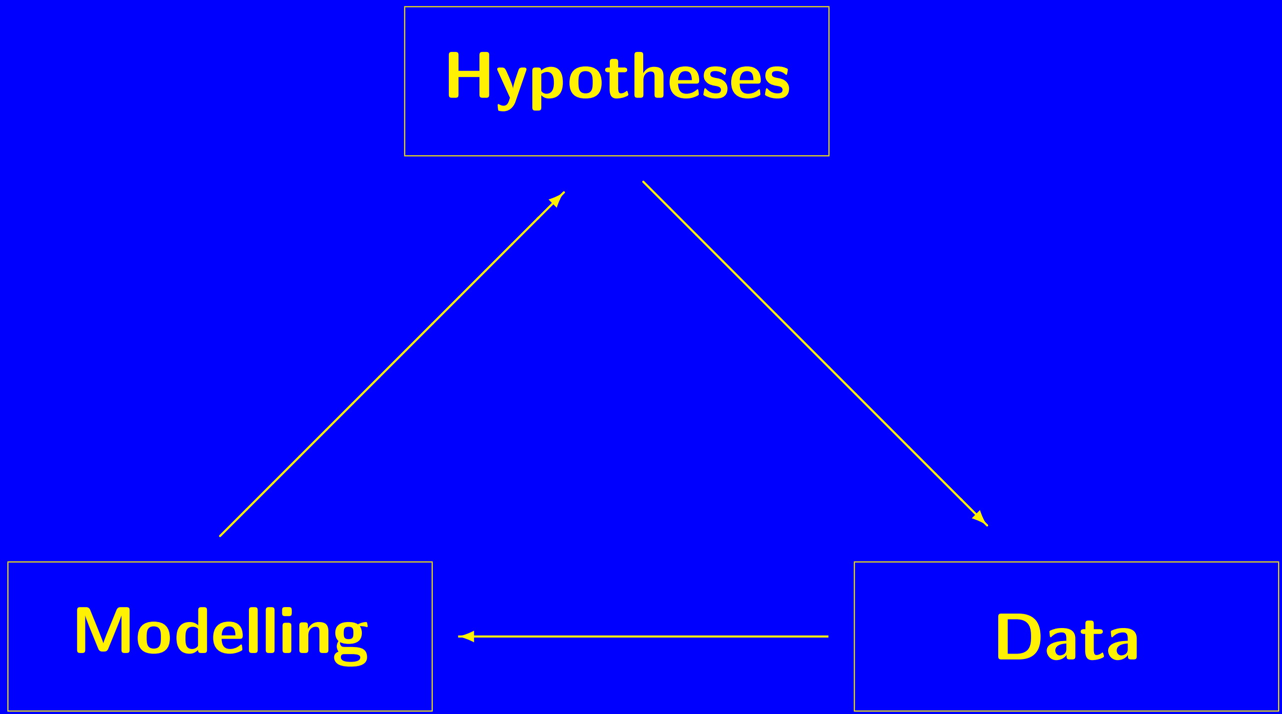
Where Do The Parameters Come From ?

Canonical form of models:

$$\dot{\vec{x}} = \vec{f}(\vec{x}, \vec{p}, \vec{u})$$

- **Function $\vec{f}(.)$ from pathways cartoon**
- **Input $\vec{u}(t)$ measured**
- **Parameters $\vec{p}$:**
  - **"Taken from the literature"**
    Problem: Different conditions, cell systems, ...
  - **Estimated from time-resolved, quantitative data**
    Poses new challenges

The Systems Biology Cycle: A Process



$$\dot{\vec{x}} = \vec{f}(\vec{x}, \vec{p}, \vec{u}) \quad \textbf{Dynamics} \quad \vec{x} \in \mathbb{R}^n_+$$

$$\vec{y}(t_i) = \vec{g}(\vec{x}(t_i), \vec{p}) \quad \textbf{Observations} \quad \vec{y} \in \mathbb{R}^m_+$$

Parameter Estimation in Nonlinear Partially Observed Noisy Dynamical Systems

Dynamics:

$$\dot{\vec{x}} = \vec{f}(\vec{x}, \vec{p}, \vec{u})$$

Observations:

$$\vec{y}(t_i) = \vec{g}(\vec{x}(t_i), \vec{p}) + \vec{\epsilon}(t_i), \qquad \vec{\epsilon}(t_i) \sim N(0, \Sigma_i)$$

Log-Likelihood:

$$\chi^2(\vec{p}, \vec{x}(t_0)) = \sum_{i=1}^{N} \sum_{j=1}^{M} \left(\frac{(y_j^D(t_i) - g_j(\vec{x}(t_i; \vec{p}, \vec{x}(t_0))}{\sigma_{i\,j}} \right)^2$$

Structural (Non-)Identifiability: Trivial Example

- **Consider:** $y = a\, e^{b+cx} = a\, e^{b}\, e^{cx}$
- **If fitted to data, only**

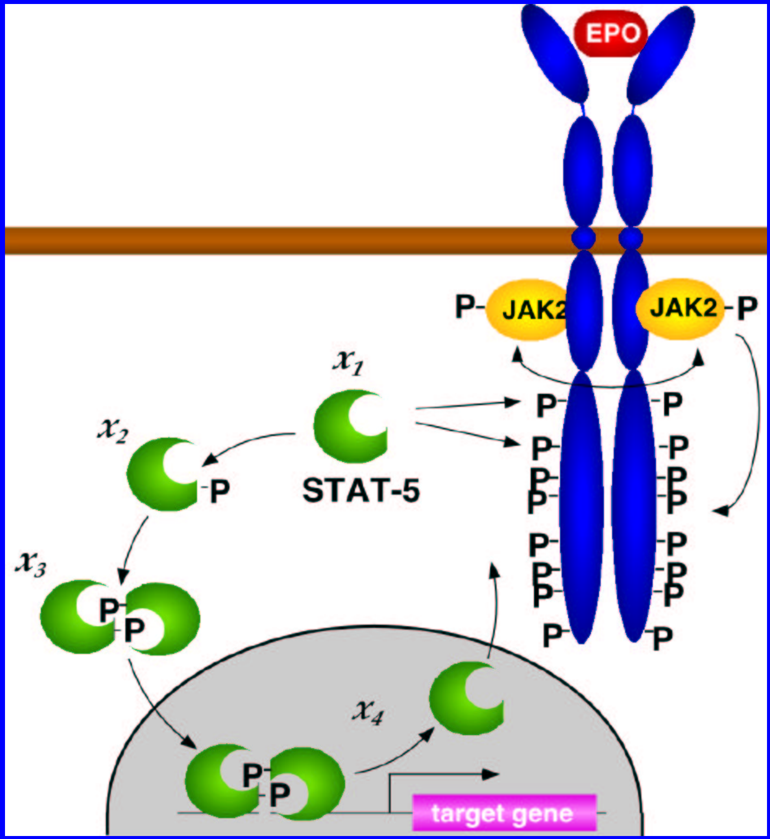
$$d = a\, e^{b}$$

  can be determined, neither a nor b individually

- **Relationship between non-ident. parameters:** $a = d\, e^{-b}$
- $\chi^2(\vec{p}) = const$ **for that relationship**

Practical non-identifiability:
Large confidence intervals due to poor data quality

Structural Identifiability: Non-Trivial Example



Swameye et al. PNAS 100, 2003, 1028-1033

Structural Identifiability: Non-Trivial Example

$$\begin{aligned} \dot{x}_1 &= 2p_4x_3^\tau - p_1x_1EpoR_A & y_1(t_i) &= p_5\,EpoR_A(t_i) \\ \dot{x}_2 &= p_1x_1EpoR_A - p_2x_2^2 & y_2(t_i) &= p_6(x_2(t_i) + 2\,x_3(t_i)) \\ \dot{x}_3 &= \frac{1}{2}p_2x_2^2 - p_3x_3 & y_3(t_i) &= p_7(x_1(t_i) + x_2(t_i) + 2\,x_3(t_i)) \\ \dot{x}_4 &= p_3x_3 - p_4x_3^\tau \end{aligned}$$

Non-identifiable pairs:

$$p_2\,x_1(0),\;\; p_1/p_5,\;\; p_6/p_2,\;\; p_7/p_2$$

Structural Identifiability: The Problem

Given:

$$\begin{aligned}\dot{\vec{x}} &= \vec{f}(\vec{x}, \vec{p}, \vec{u}) \quad &\textbf{Dynamics}\\ \vec{y}(t_i) &= \vec{g}(\vec{x}(t_i), \vec{p}) \quad &\textbf{Observations}\end{aligned}$$

Question:

- **Given** $\{\vec{u}, \vec{f}(.), \vec{g}(.), t_i\}$**, can** $\vec{p}$ **be uniquely determined ?**

Existing methods:

- **Analytical approaches: Only applicable to small systems**
- **Approximative methods: Hardly controllable**

Non-Identifiability and Systems Analysis

- **The model in itself is not the goal**
- **Goal: Systems analysis based on the model**

Consequences of non-identifiability for systems analysis:

- **Confidence intervals for identifiable parameters: possible**
- **Summation theorems: Not affected**
- **Predictions and extrapolations: It depends**

  Non-identifiability is coupled to non-observability

(Non-)Observability

Given:

$$\begin{aligned} \dot{\vec{x}} &= \vec{f}(\vec{x}, \vec{p}, \vec{u}) && \textbf{Dynamics} \\ \vec{y}(t_i) &= \vec{g}(\vec{x}(t_i), \vec{p}) && \textbf{Observations} \end{aligned}$$

Question:

- **Given $\{\vec{u}, \vec{f}(.), \vec{g}(.), t_i\}$, can $\vec{x}(t)$ be uniquely determined ?**

If some p_i are <u>non-identifiable</u>

$$\Longrightarrow$$

Some $x_i(t)$ will be <u>non-observable</u>

Approximative Methods

- **Structural non-identifiability:**

  $\exists$ **continuous set of parameters with constant** $\chi^2(p)$

- **Consider curvature** H **of** $\chi^2(\hat{\vec{p}})$

$$H = \frac{\partial^2 \chi^2(\hat{\vec{p}})}{\partial p_i \, \partial p_j},$$

Asymp. confidence intervals from H^{-1}

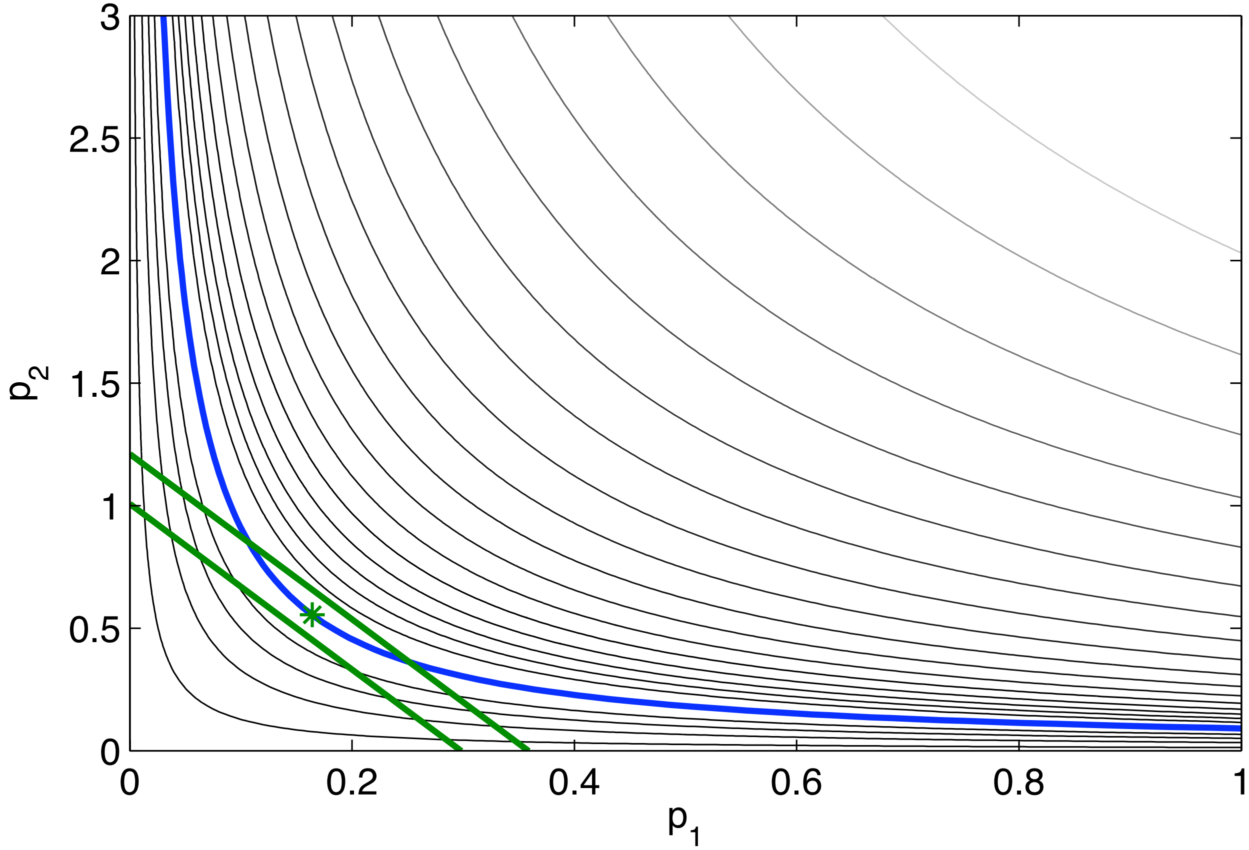
- **Evaluate eigen-values of** H**:**

  Non-identifiabilities should correspond to zero eigen-values

- **Problem: Non-linearity of the parameter relationships**

Approximative Methods: Example

χ^2**-landscape, non-identifiability:** $p_1\, p_2 = const$



The Idea of the New Method

Structural non-identifiability:

- **Functional relationships between parameters**
- $\chi^2(\vec{p})$ **does not change along these relationships**

Idea: Do changes of $\hat{\vec{p}}$ **exist that do not change** $\chi^2(\vec{p})$ **?**

Profile Likelihood and Confidence Regions

- **Profile likelihood:**

$$PL_i : \chi^2(p_i) = \min_{p_{j \neq i}} [\chi^2(\vec{p})]$$

Likelihood of p_i with all other parameters re-optimized

- **Confidence regions determined by increase of likelihood**

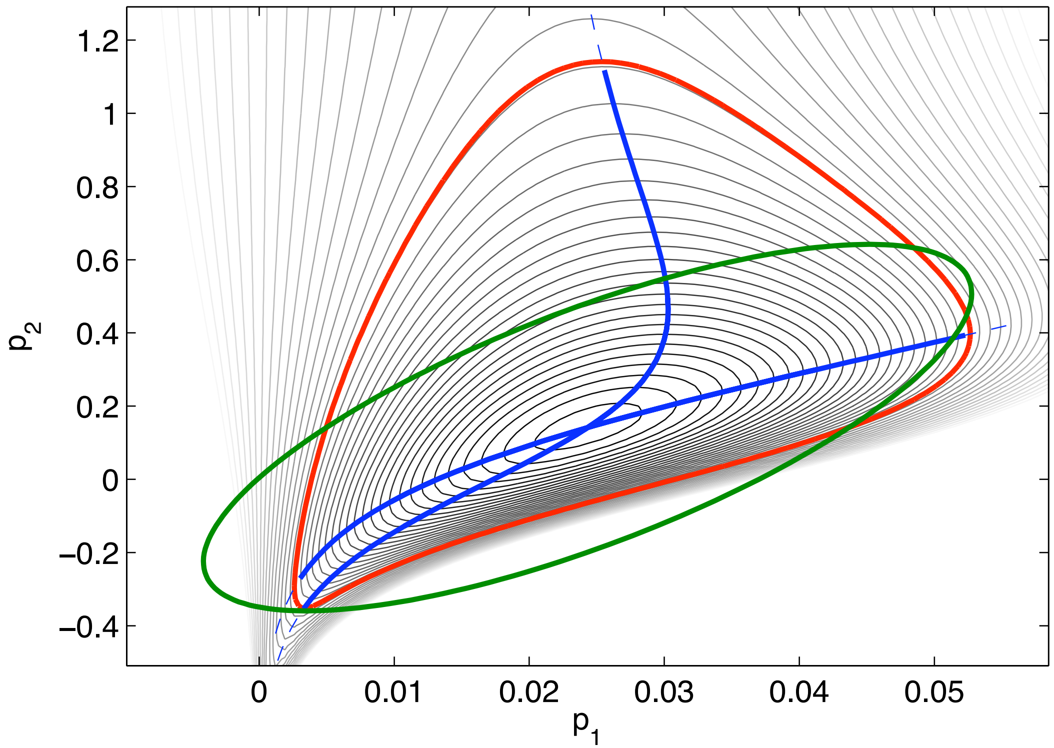
$$\chi^2(\vec{p}) - \chi^2(\hat{\vec{p}}) < \chi^2_{(1-\alpha,r)}$$

$r = 1$ **pointwise,** $r = \#p$ **simultaneous confidence regions**

Confidence Regions and Profile Likelihood

χ^2-landscape

Asymp. CR **Likelihood CR** **Profile likelihood**



Structural and Practical Identifiability

Consider threshold $\Theta = \chi^2(\hat{\vec{p}}) + \chi^2_{(1-\alpha,r)}$

- **Structural and practical identifiable:**
  - PL_i **crosses** Θ **for** $\hat{p}_i - \sigma_-$ **and** $\hat{p}_i + \sigma_+$
  - $[\hat{p}_i - \sigma_-, \hat{p}_i + \sigma_+]$ **represent confidence intervals**

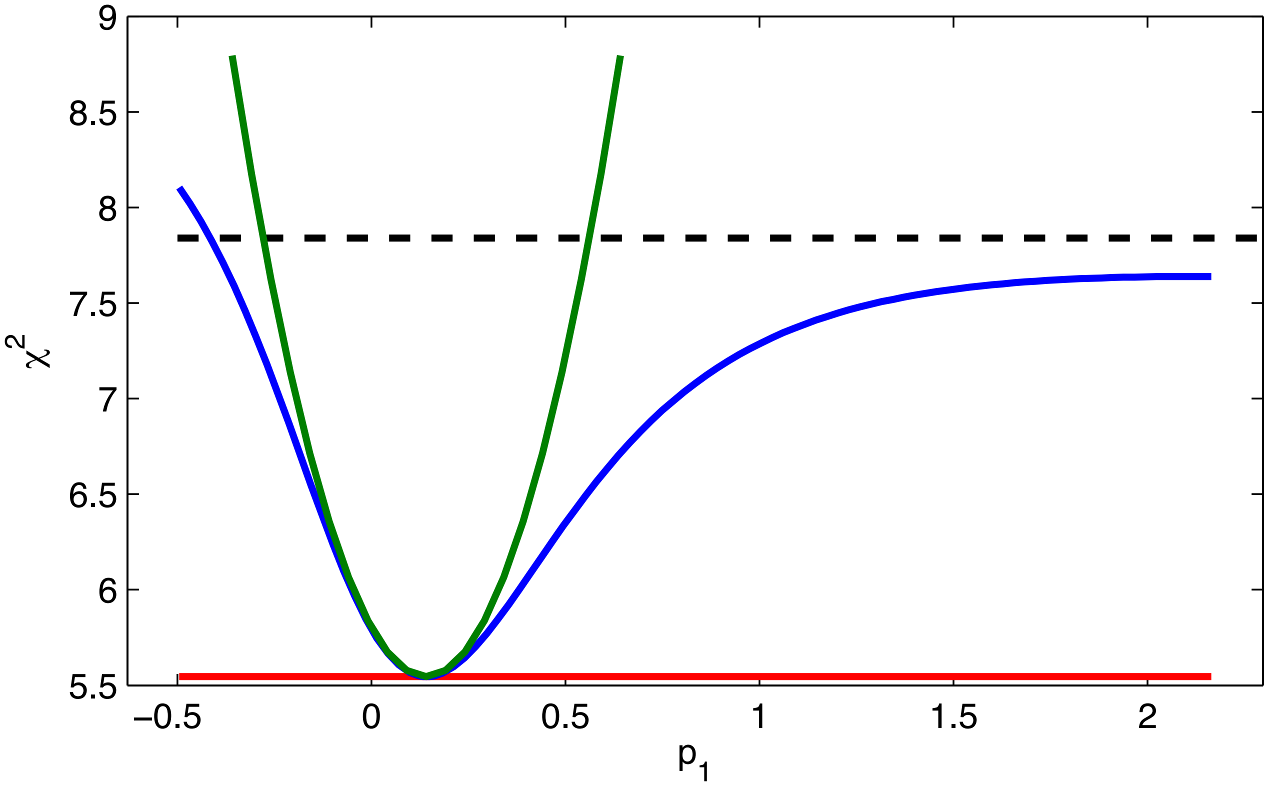
- **Structural non-identifiable:** $PL_i = const$

- **Practical non-identifiable:**

  $PL_i \neq const$, σ_+ **and/or** σ_- **=** ∞ **(on log. scale)**

The Three Cases

identifiable structural non-identifiable practical non-identifiable

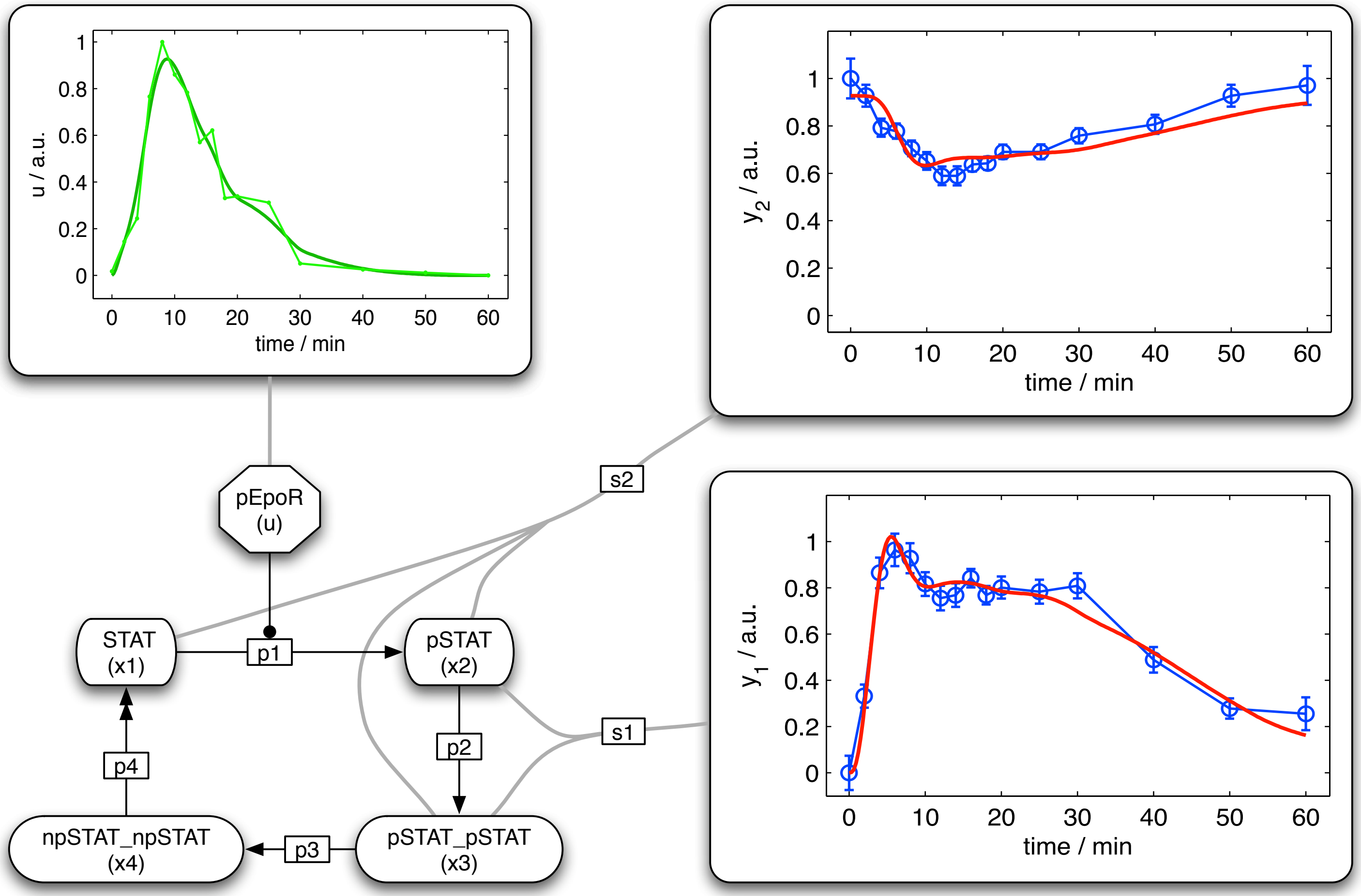


Find Functional Relationships

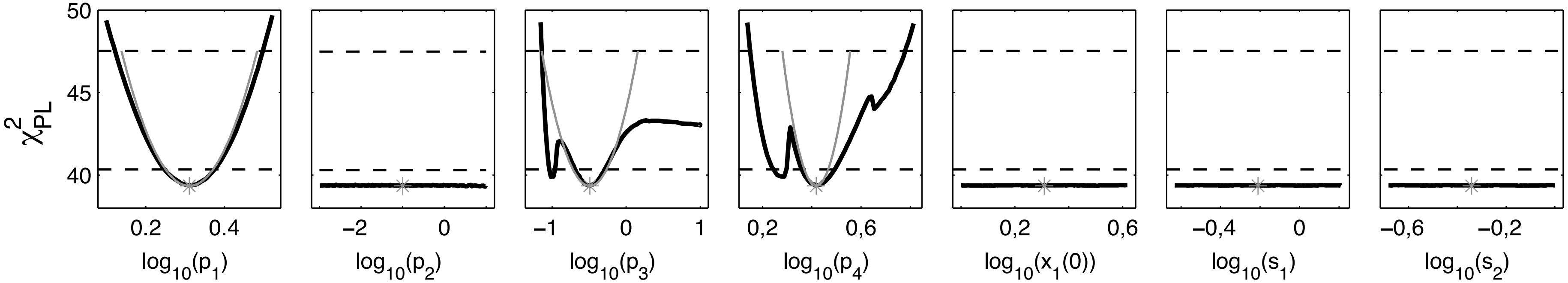
If one non-identifiable parameter p_i is identified:

- **Plot all other parameters in dependence of p_i**

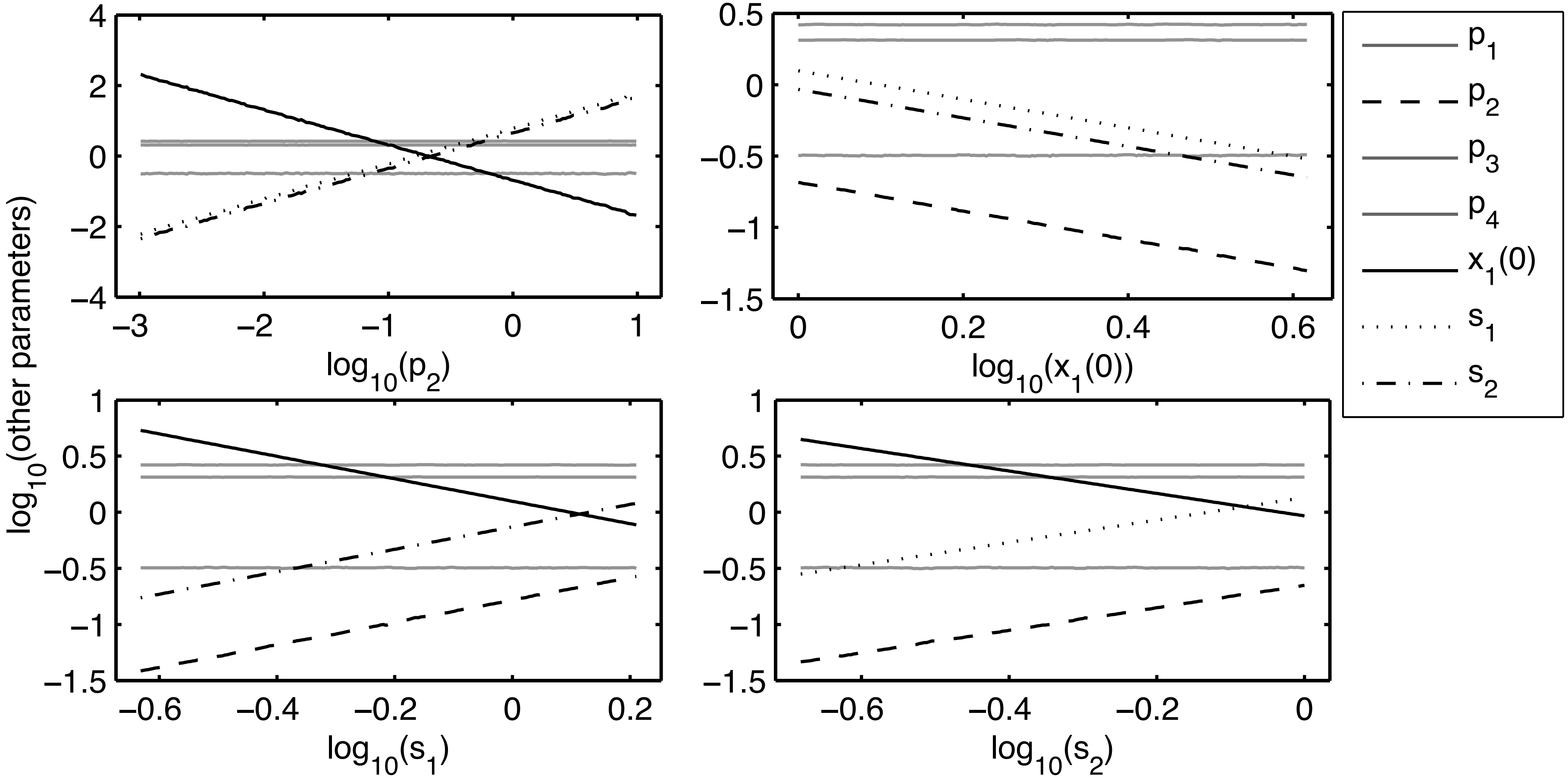
An Example: JAK-STAT pathway



Profile Likelihood

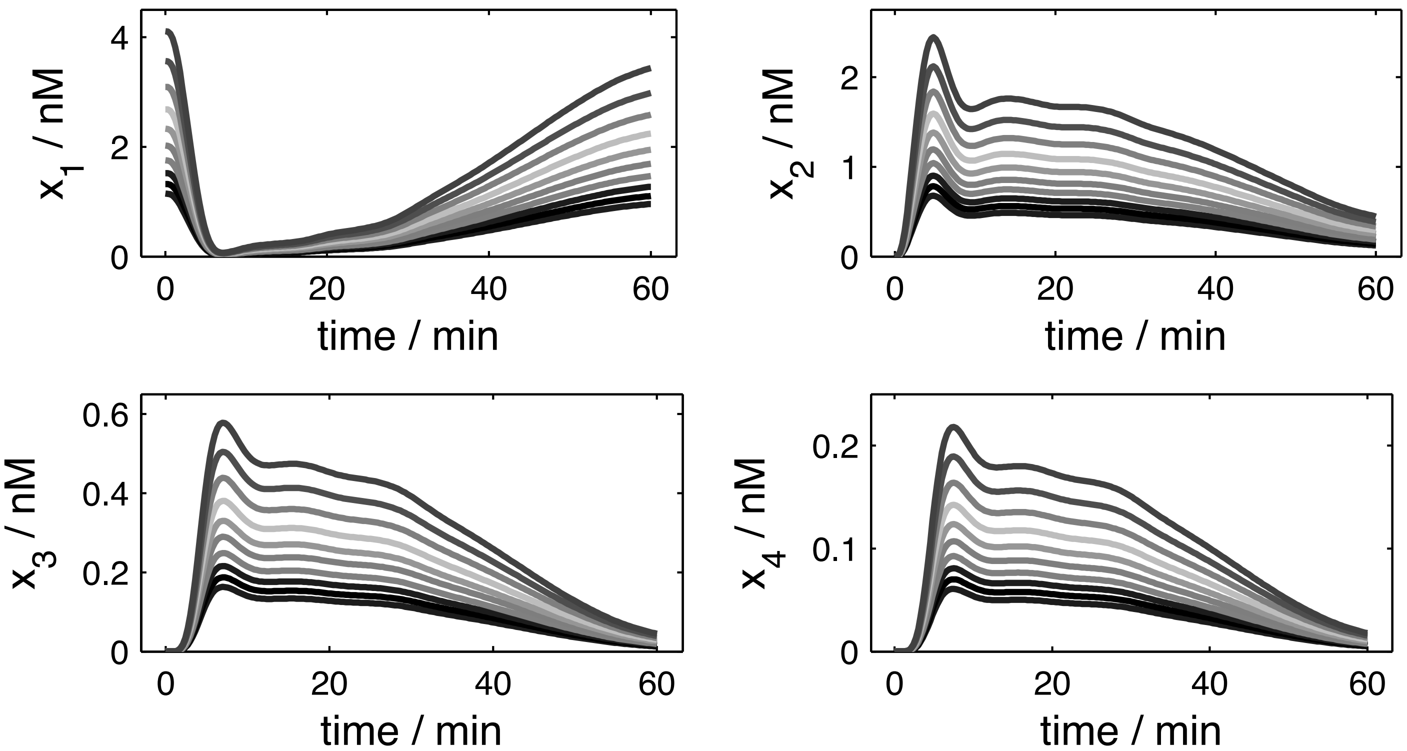


Relations of Non-Identifiable Parameters



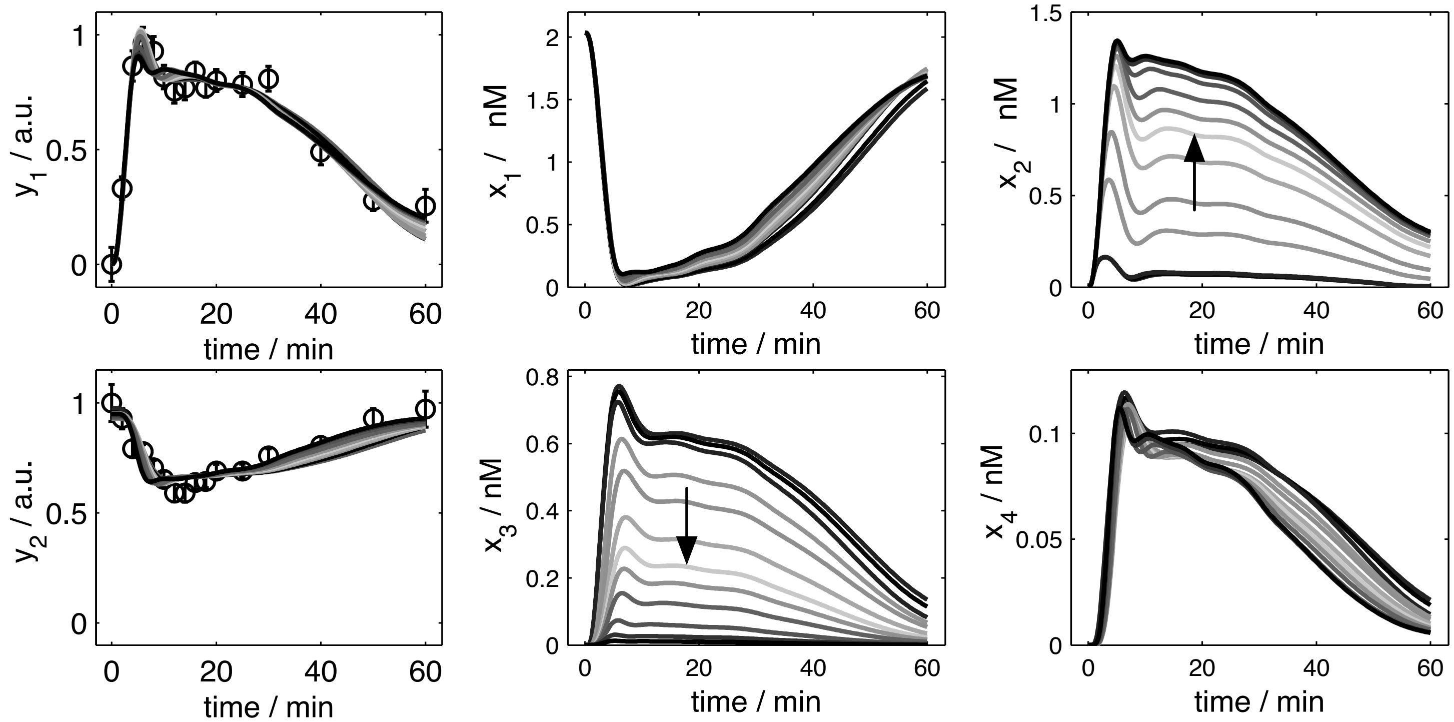
Non-Observability

Non-observability due to structural non-identifiability



Non-Observability

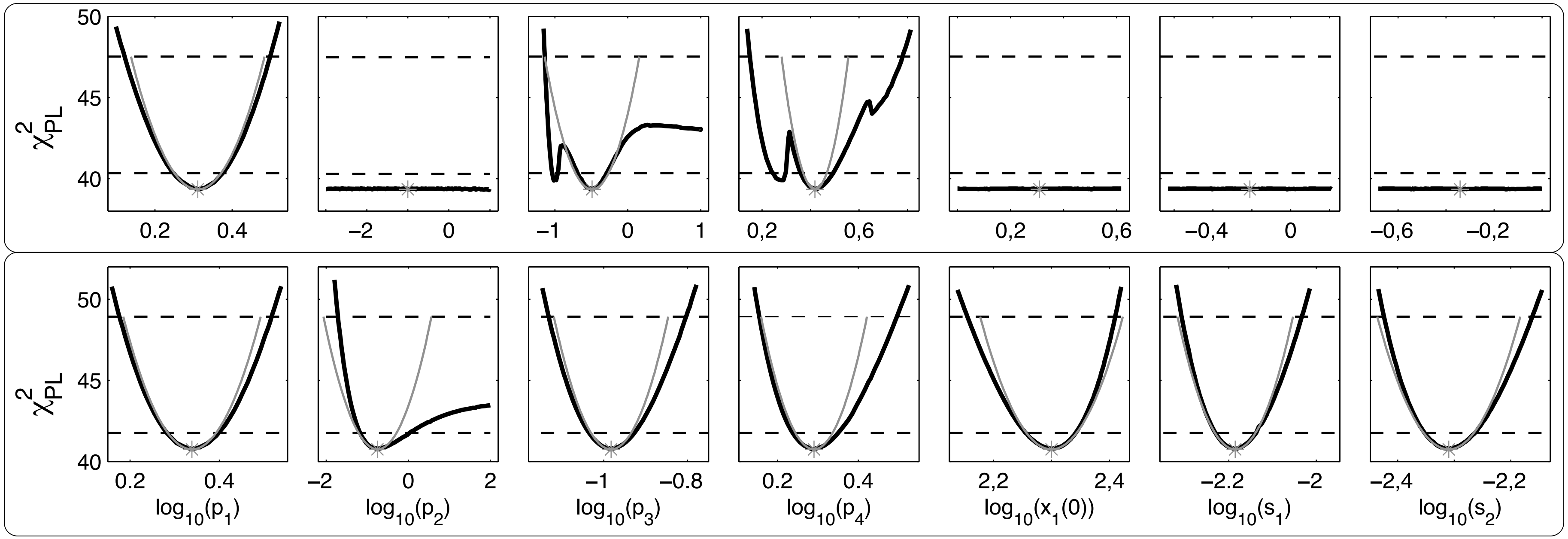
Non-observability due to practical non-identifiability of p_3



Experimental Design

Observability analysis suggests two additional measurements

- $x_1(0) = 200 \pm 20nM$
- $x_3/(x_2 + x_3) = 0.9 \pm 0.05nM \;\; at \;\; t = 20 \;\; min$



Properties of the Method

- No assumptions about functional form of non-identifability
- Applicable to large systems
- Applicable to any kind of parameter estimation problem
  - Ordinary differential equations
  - Stochastic differential equations
  - Partial differential equations
  - Any continuous parameter estimation problem

Benefit

- Experimental design: What to measure when ?
- Model reduction: Lump processes/parameters

Goals:

- Tailor model complexity to information content of data
- Turn all parameters identifiable
- Turn all experimentally unobserved components observable
- Obtain reliable model predictions

Papers and Software

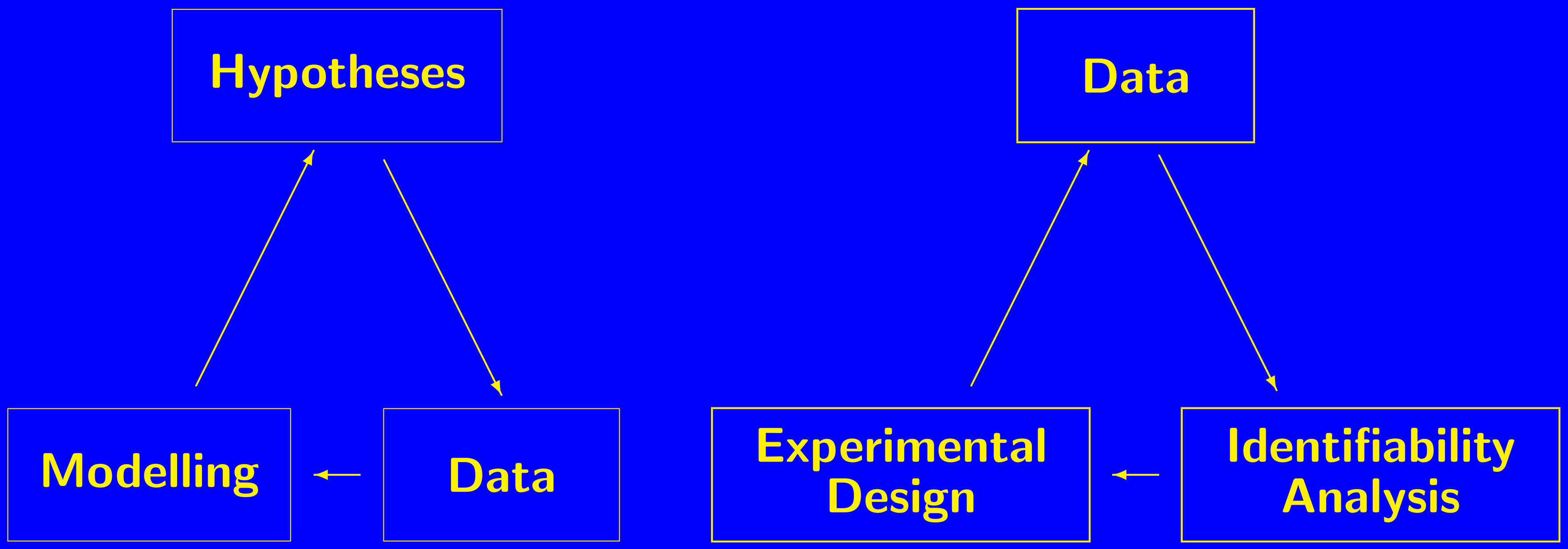
A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, J. Timmer
Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics, 25, 2009, 1923-1929

Hengl S., Kreutz C., Timmer J. Maiwald T
Data-dased identifiability analysis of nonlinear dynamical models. Bioinformatics 23, 2007, 2612-2618

Both methods are included in modelling software *PottersWheel*: **www.potterswheel.de**

T. Maiwald, J. Timmer
Dynamical modeling and multi-experiment fitting with PottersWheel. Bioinformatics 24, 2008, 2037-2043

Summary: The Two Systems Biology Cycles



Acknowledgements

Theoretical side	Experimental side DKFZ, Heidelberg
Andreas Raue	Verena Becker
Thomas Maiwald	Marcel Schilling
Clemens Kreutz	Julie Bachmann
	Ursula Klingmüller

SBMC 2010

3rd Conference “Systems Biology of Mammalian Cells”

June 3-5, 2010

Freiburg, Germany

www.sbmc2010.de