Data-based modeling of the dynamics of a cellular signaling pathway

Jens Timmer

In cooperation with:
Klingmüller Lab, DKFZ

Center for Data Analysis and Modeling
Center for Systems Biology
Department for Mathematics and Physics
University of Freiburg

http://www.fdm.uni-freiburg.de/~jeti/
Outline

• Introduction

• JAK-STAT pathway of the Epo receptor

• Simulation and data based modeling

• A dynamical model for JAK-STAT pathway

• Observing the unobservable

• In silico biology: Predicting a new experiment

• Outlook
Man: A Dynamical System

Diseases caused or expressed by malfunction of dynamical processes
General Goal

Understand biomedical systems by data-based analysis of their dynamical behavoir.

Time Series Analysis
Two Gaps In Time Series Analysis

Specific questions

Time series

Time series analysis

Specific applications
Two Gaps In Time Series Analysis

- Specific questions
- Time series analysis
- Time series

- Time series analysis
- Specific questions
- Specific applications
Methods

Time-frequency-distributions

linear stochastic processes

Parameter fitting in ODEs

measured time series

Hidden-Markov-Models

nonlinear deterministic processes

Statistics and Numerics
Goals of Time Series Analysis

- Prediction
- Characterization / Classification
  - Improvement of diagnosis and therapy
- Modeling
  - Hypotheses testing
  - Understanding
  - Control
Why Modelling in BioMed?

- Make assumptions explicit
- Understand essential properties, failing models
- Condense information, handle complexity
- Understand role of dynamical processes, e.g. feedback
- Prediction and control
- Discover general principles
- "You don't understand it until you can model it"
Why Modelling in Cell Biology?

- ...omics does not necessarily lead to understanding of function
- Function determined by regulation
- Regulation = Dynamics
- Function: Property of dynamic network
- "Systems Biology"
Signal transduction through the Erythropoietin receptor (EpoR)
Mass Action Yields:

\[
\begin{align*}
\dot{x}_1 &= -p_1 x_1 EpoR_A \\
\dot{x}_2 &= p_1 x_1 EpoR_A - p_2 x_2^2 \\
\dot{x}_3 &= \frac{1}{2} p_2 x_2^2 - p_3 x_3 \\
\dot{x}_4 &= p_3 x_3
\end{align*}
\]
Measurements

- $y_1(t)$: Phosphorylated STAT-5 in the cytoplasm
  \[ y_1(t) = p_5(x_2(t) + 2x_3(t)) \]

- $y_2(t)$: All STAT-5 in the cytoplasm
  \[ y_2(t) = p_6(x_1(t) + x_2(t) + 2x_3(t)) \]

- $y_3(t)$: Activation of the epo receptor
  \[ y_3(t) = p_7 \text{EpoR}_A(t) \]
Simulation vs. Data Based Modeling 1

Model comprises:

- **Structure of the equations (the cartoon)**
- **Values of the parameters**

Simulation:

- **Structure from pathway cartoon**
- **Parameters from**
  - Independent measurements
  - Literature
  - Educated guesses
Simulation vs. Data Based Modeling II

Simulation dilemma:

If discrepancies between experiment and model

- Wrong structure or wrong parameters?

Data based modeling:

- Structure from pathway cartoon
- Parameters estimated from data

If discrepancies:
  Think about the cartoon!  Learn biology!
Parameter Estimation

Dynamics:

\[ \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{p}) \]

Observation:

\[ \tilde{\mathbf{y}}(t_i) = \mathbf{g}(\mathbf{x}(t_i), \mathbf{p}) + \epsilon(t_i) \quad \epsilon(t_i) \sim N(0, \Sigma_i) \]

Log-Likelihood:

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

Initial Value Approach, Multiple Shooting, GO
Statistics I

• Confidence regions for parameters
  – Asymptotically:
    \[ \frac{\partial^2}{\partial p_i \partial p_j} \chi^2(\hat{p}, \hat{x}(t_0)) \]
  – Finite:
    * Log-Likelihood contours
    * Bootstrap
• Model selection
  – Likelihood ratio test
  – Non-standard test situations:
    * Parameter on the boundary
    * Non-identifiability under the null
  – Non-nested models, Bootstrap
"What makes you feel good?"

"Good data."

"What makes you feel really good?"

"Really good data!"
Quantitative Immunoblotting

<table>
<thead>
<tr>
<th>Immunoprecipitation:</th>
<th>Immunoblotting:</th>
</tr>
</thead>
<tbody>
<tr>
<td>anti-EpoR</td>
<td>anti-PTyr</td>
</tr>
<tr>
<td>anti-STAT-5</td>
<td>anti-PTyr</td>
</tr>
<tr>
<td>anti-STAT-5</td>
<td>anti-STAT-5</td>
</tr>
</tbody>
</table>

Epo  
\[ \text{pY-EpoR} \]

\[ \text{pY-STAT-5} \]

\[ \text{STAT-5} \]

0 2 4 6 8 10 12 14 16 18 20 25 30 40 50 60 min
Really Good Data

\[ g(x) \text{ is linear} \]
The data

Activation of the epo receptor:

Maximum at 8 min
The data

Phosphorylated STAT-5 in cytoplasm:

Plateau from 10 to 30 min
The data

All STAT-5 in cytoplasm:

![Graph showing STAT-5 in cytoplasm over time](image-url)
First results

Phosphorylated STAT-5 in cytoplasm:

![Graph showing the time course of phosphorylated STAT-5 in the cytoplasm with error bars and a trend line.](image_url)
First results

All STAT-5 in cytoplasm:

![Graph showing the total STAT-5 in cytoplasm over time.]
JAK – STAT Pathway
Model Extension
Modeling Nuclear Export

- One compartment:
  \[ \dot{x}_4 = p_3x_3 - p_4x_4 \]

- Series of compartments \( \approx \) delay
  \[ \dot{x}_4 = p_3x_3 - p_4x_3^\tau, \quad x_3^\tau = x_3(t - \tau) \]

- Non-nested models
\[\dot{x}_1 = 2p_4 x_3^\tau - p_1 x_1 \text{EpoR}_A\]
\[\dot{x}_2 = p_1 x_1 \text{EpoR}_A - p_2 x_2^2\]
\[\dot{x}_3 = \frac{1}{2} p_2 x_2^2 - p_3 x_3\]
\[\dot{x}_4 = p_3 x_3 - p_4 x_3^\tau\]
Results

Phosphorylated STAT-5 in cytoplasm:

$\tau \approx 6 \text{ min}$
All STAT-5 in cytoplasm:
Observing the unobservable

Simulating the fitted model:
Access to dynamic variables $x_i$

- Unphosphorylated STAT-5 is limiting factor

- Experimental fact:
  Phosphorylated monomeric STAT-5 is hard to measure

Explanation by the model:
It is rapidly processed into dimeric STAT-5
In silico biology

What happens if ... ?

- Sensitivity analysis
- Change parameters in the model
- Calculate the transsscriptional yield
Prediction of New Experiment

- Result of sensitivity analysis:
  
  Transcriptional yield is most sensitive to nuclear shuttling parameters.

- Setting $p_4 = 0$ or $\tau = \infty$
  
  Only one cycle: Only 45% efficiency

- Blocking nuclear export by leptomycin B confirms prediction.

Perspective:

Identification of targets for medical intervention
Experimental Confirmation of Prediction
Why Cycling?

- Optimal use of limited pool of STAT-5
- Continuous monitoring of receptor activity:

  Systems’ property: ”Remote Sensor”

"All models are wrong ..."

- No scaffolding for receptor–STAT-5 interaction, 200 eqs.
- Spatial effects, ODE vs. PDE
- Stochastic effects

"... but some are useful"

- Captures the main effect
- Makes testable prediction

Successful modelling: Making controlled "errors"
Summary

Given time-resolved "really good" data:

It is possible to turn qualitative cartoons into quantitative dynamical models

- Testing the cartoon
- Calculating unobservable components
- Manipulating the system \textit{in silico}
- Prediction of experiments
- Infering systems’ properties
Outlook: Scale It Up

BMBF Systems Biology of (Regenerating) Hepatocytes

- SMAD, IGF, Wnt/β-catenin, NF-κB, … pathways
- Crosstalk
- Interaction Transcription factors – DNA
- Genetic Networks
The Mission of Systems Biology

**Turn the life sciences**

*from a*

**static, qualitative, descriptive**

*into a*

**dynamic, quantitative, predictive science**