Quantification of Metabolic Pathway Models: Beyond Acceptable Parameter Fits

Eberhard O. Voit
Department of Biomedical Engineering
Georgia Institute of Technology and Emory University

Workshop on Parameter Estimation for Dynamical Systems
4-6 June 2012, Eurandom, Eindhoven, The Netherlands

Overview

Nederlandse Spoorwegen: Advice on Parameter Estimation

Geloof in wonderen, maar ben er niet afhankelijk van.

Believe in miracles, but better not depend on them.

Overview

Introduction

Caveat emptor: nothing stochastic here!

Generic Issues of Parameter Estimation

Estimation Strategies: “From processes up” versus “From time series down”

Beyond Quality of Fit

System Identification

“Non-parametric” Dynamic Flux Estimation (DFE)

Challenges of DFE and Partial Remedies

Definitions

Parameter:

A quantity in a function or set of equations that remains constant during a mathematical evaluation ("computational experiment"), but may vary from one experiment to the next.

Parameter Estimation (Mathematics):

The process of identifying values of parameters in a model that (typically) minimize the difference between the output of the model and corresponding data.

Example:

$$F(x) = m\,x + b$$

Overall Goal

Parameter estimation in systems analysis requires that we know the functional form of the model or set of equations.

In contrast to statistics, there seem to be no widely-accepted “nonparametric methods” in dynamical systems modeling (outside analog modeling; Ellner et al. 2002 used spline regression).

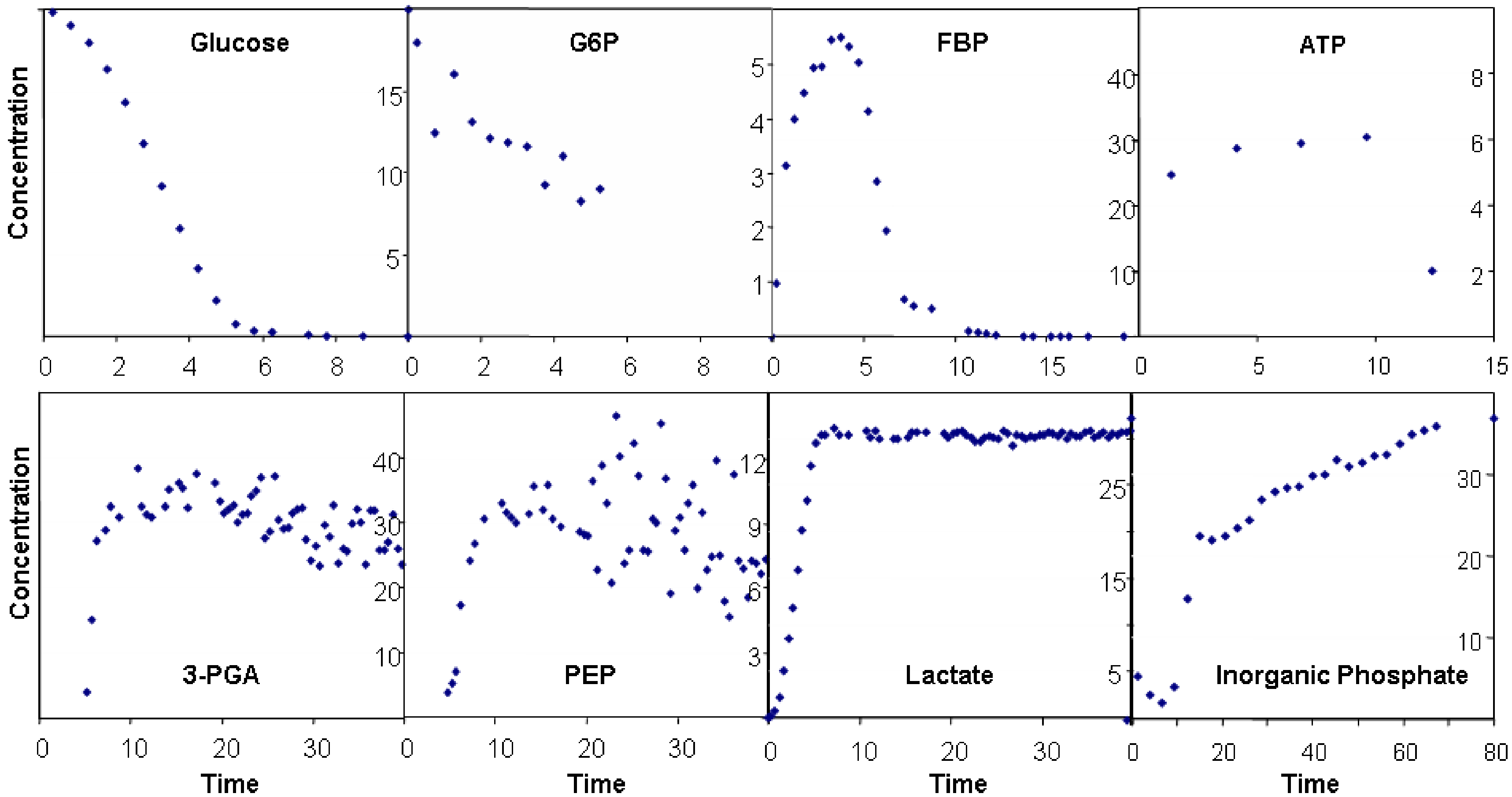
Goal here: slightly ameliorate the problem (without completely solving it)

Diagnostics of Core Problem: Why don't we have functions?

Physics: Functions come from theory

Biology: No theory available

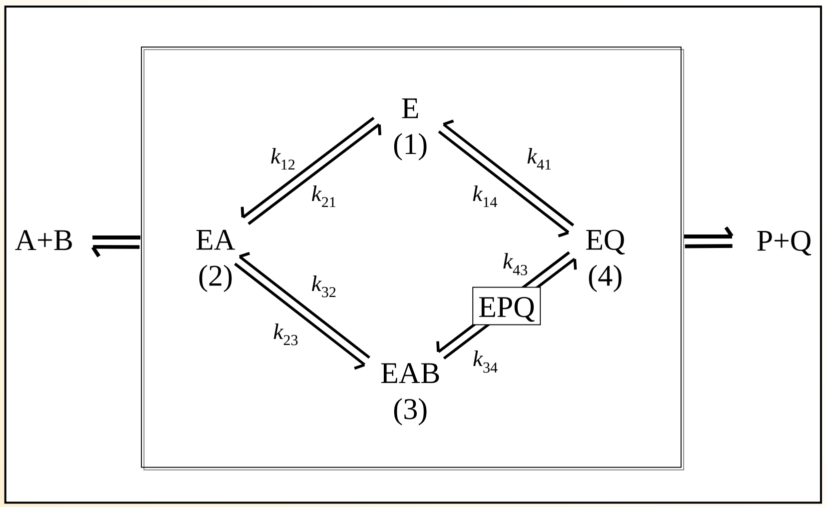
Example: Glycolysis in Lactococcus



Voit *et al.*: *IEE Proc. Systems Biol. 2006; PNAS 2006*

Why Not Use "True" Rate Functions?

$$\text{A+B} \rightleftharpoons \text{P+Q}$$



$$v = \frac{\left(\frac{\text{num.1}}{\text{coef. AB}}\right)(\text{A})(\text{B}) - \left(\frac{\text{num.1}}{\text{coef. AB}} \times \frac{\text{num.2}}{\text{num.1}}\right)(\text{P})(\text{Q})}{\left(\frac{\text{constant}}{\text{coef. A}} \times \frac{\text{coef. A}}{\text{coef. AB}}\right) + \left(\frac{\text{coef. A}}{\text{coef. AB}}\right)(\text{A}) + \left(\frac{\text{coef. B}}{\text{coef. AB}}\right)(\text{B}) + \left(\frac{\text{coef. AB}}{\text{coef. AB}}\right)(\text{A})(\text{B}) + \left(\frac{\text{coef.P}}{\text{coef. AP}} \times \frac{\text{coef. AP}}{\text{coef. A}} \times \frac{\text{coef. A}}{\text{coef. AB}}\right)(\text{P}) + \left(\frac{\text{coef.Q}}{\text{constant}} \times \frac{\text{constant}}{\text{coef. A}} \times \frac{\text{coef. A}}{\text{coef. AB}}\right)(\text{Q}) + \left(\frac{\text{coef.AP}}{\text{coef. A}} \times \frac{\text{coef.A}}{\text{coef. AB}}\right)(\text{A})(\text{P}) + \left(\frac{\text{coef.BQ}}{\text{coef. B}} \times \frac{\text{coef.B}}{\text{coef. AB}}\right)(\text{B})(\text{Q}) + \left(\frac{\text{coef.PQ}}{\text{coef.Q}} \times \frac{\text{coef.Q}}{\text{constant}} \times \frac{\text{constant}}{\text{coef. A}} \times \frac{\text{coef. A}}{\text{coef. AB}}\right)(\text{P})(\text{Q}) + \left(\frac{\text{coef.ABP}}{\text{coef. AB}}\right)(\text{A})(\text{B})(\text{P}) + \left(\frac{\text{coef.BPQ}}{\text{coef. BQ}} \times \frac{\text{coef.BQ}}{\text{coef. B}} \times \frac{\text{coef.B}}{\text{coef. AB}}\right)(\text{B})(\text{P})(\text{Q})}$$

from Schultz (1994)

Diagnostics of Core Problem

Physics: Functions come from theory

Biology: No theory available

Solution 1: Educated guesses: growth functions

Solution 2: “Partial” theory: Enzyme kinetics

Solution 3: Generic approximation

Biochemical Systems Theory

$$\xrightarrow{V_i^+} \boxed{X_i} \xrightarrow{V_i^-}$$

$$\dot{X}_i = \frac{dX_i}{dt} = V_i^+ - V_i^-$$

$$V_i^+ = V_i^+(\underbrace{X_1, X_2, \ldots, X_n}_{\textit{inside}}, \underbrace{X_{n+1}, \ldots, X_{n+m}}_{\textit{outside}})$$

complicated

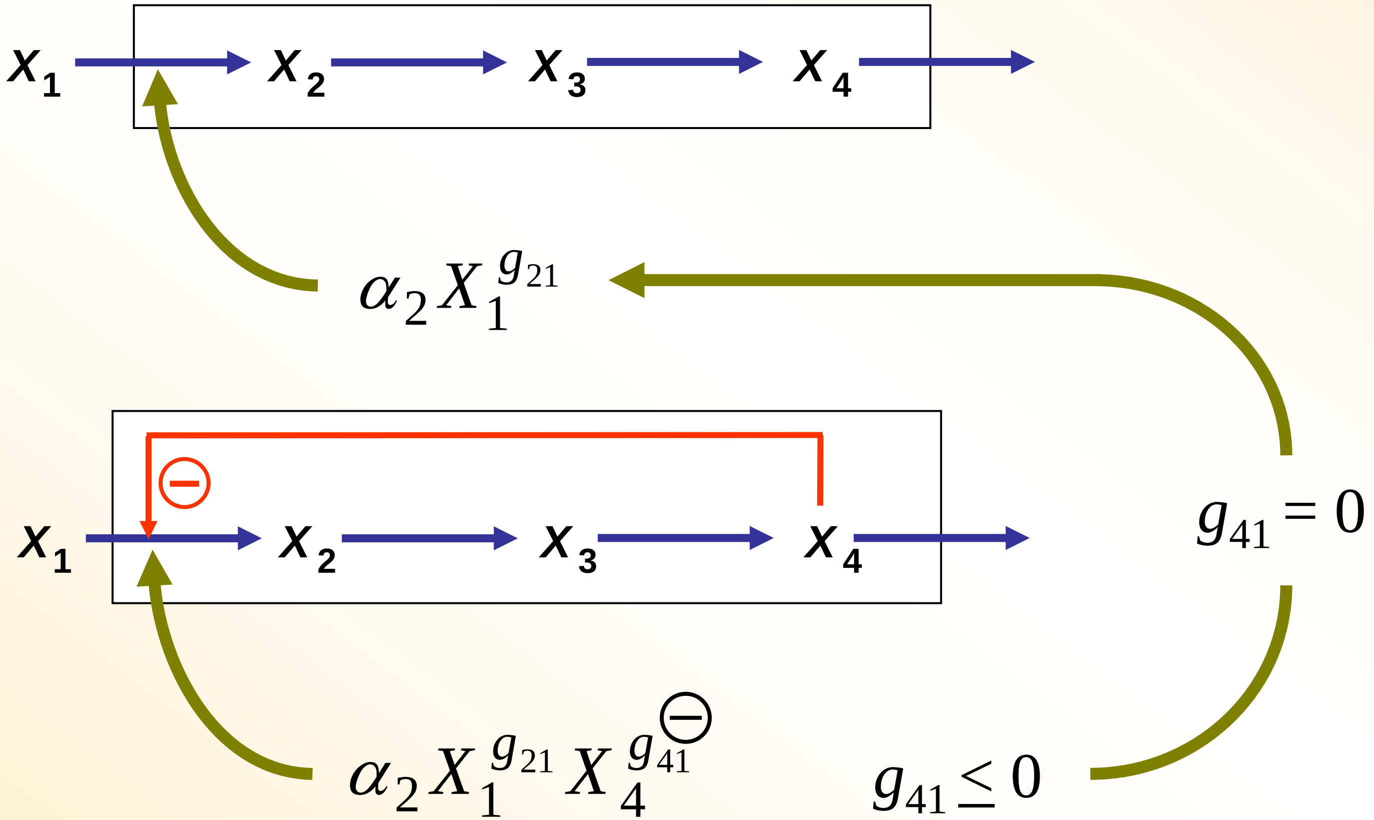
Solution with Potential:

$$\boxed{V_{ik}^{+/-} = \gamma_{i,k} \prod_{j=1}^{n} X_j^{f_{k,i,j}}}$$

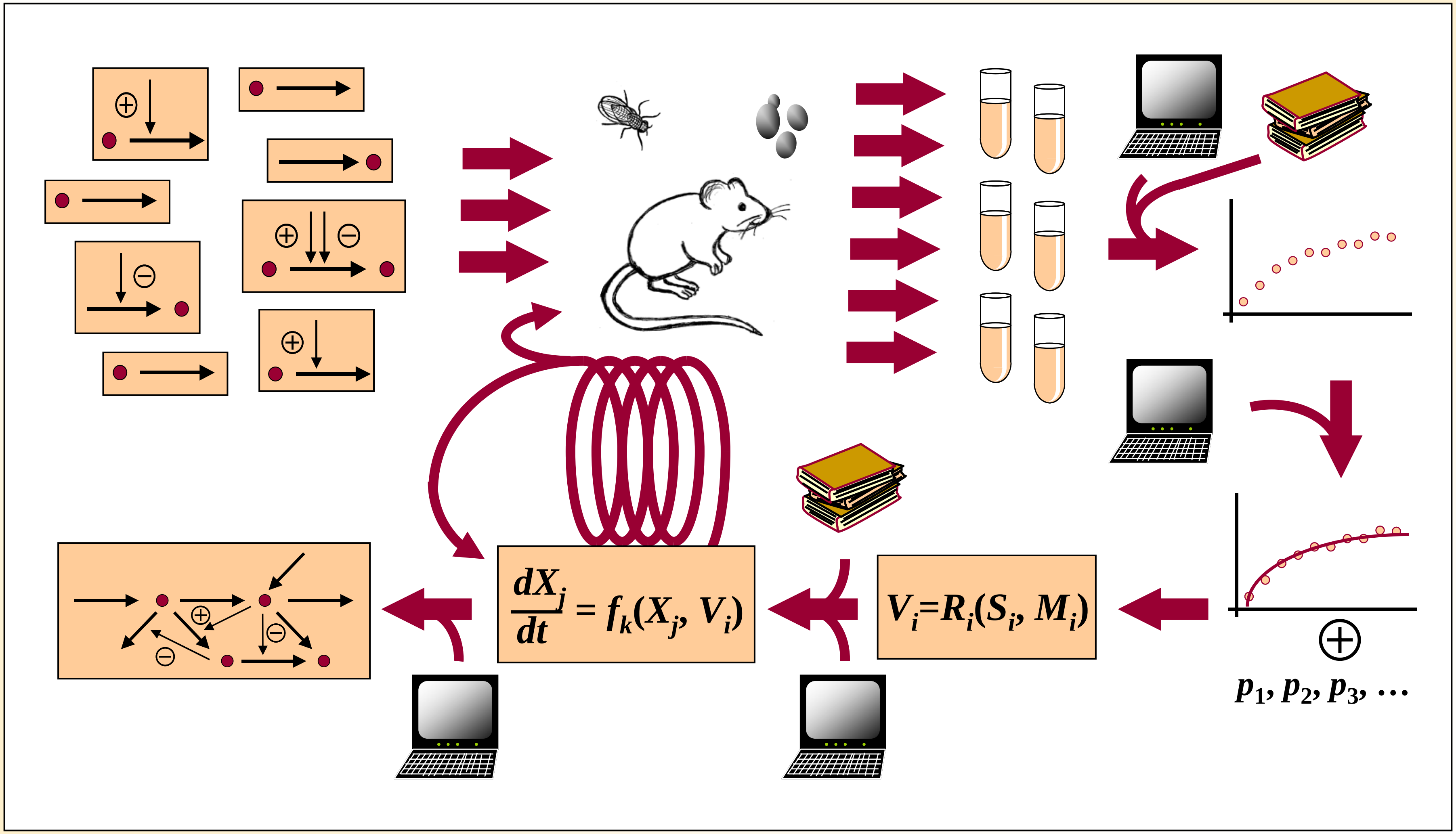
"Biochemical Systems Theory"
(BST)

Note: BST does not solve the problem of unknown functions either, but it provides a rather general and unbiased default for getting started with a model.

Mapping
Structure ⟺ Parameters

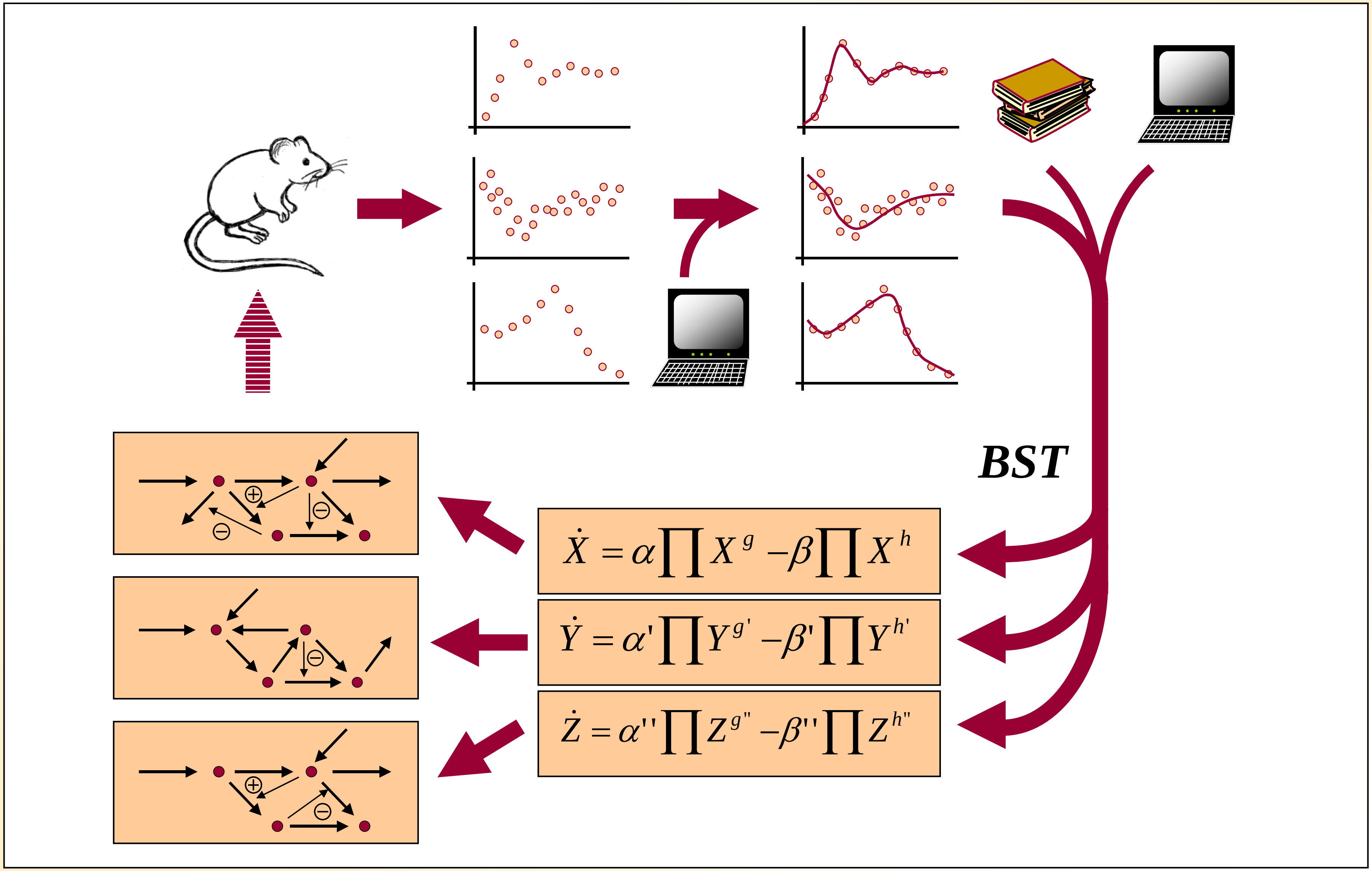


Traditional Estimation Strategy



Voit, *Drug Discovery Today*, 2004

Estimation Based on Time Series and BST



Voit, *Drug Discovery Today*, 2004

Quality of Fit

Traditional assessment of an estimation result:

Minimally possible residual error between model and data, given a fixed model structure (including a set of parameters)

Typical example: linear regression

Gutenkunst, Raue, Vilela, …:

Many almost-equivalent solutions lead to neutral spaces, sloppiness, identifiability problems.

Reasons: Too many parameters; wrong functions; too few data

One remedy: Compute ensembles of solutions, but require functional model

Vilela *et al*., *BMC Systems Biology* 3:47, 2009.

Challenges in System Estimation

Technical problems:

Time to convergence; no convergence
Very rough error surfaces
Very shallow error surfaces
Local minima

Problems with data:

Problems with collinear data
Problems with insufficient data (quantity, quality)

Problems with models:

Problems with models containing redundancies
Problems caused by similar fits with different models
Problems with compensation of error among terms

Problems with model-data combination:

Averaging of estimation results
Extrapolation

simultaneous

Challenges in System Estimation

Technical problems:

Time to convergence; no convergence
Very rough error surfaces
Very shallow error surfaces
Local minima

Problems with data:

Problems with collinear data
Problems with insufficient data (quantity, quality)

Problems with models:

Problems with models containing redundancies
Problems caused by similar fits with different models
Problems with compensation of error among terms

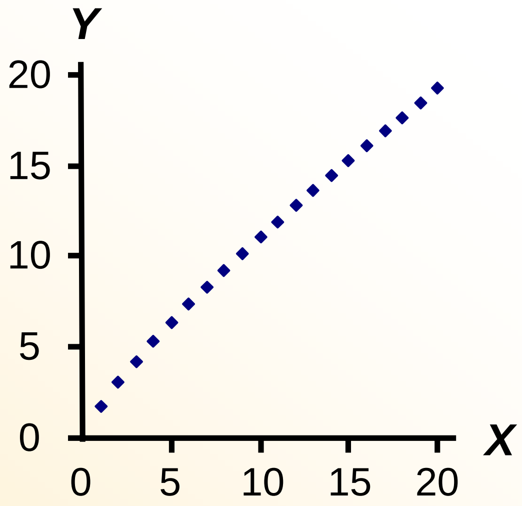
Problems with model-data combination:

Averaging of estimation results
Extrapolation

Discuss these

Problems with Model Redundancies

Example: Collinear Data (in log space):



$$Y = \alpha \cdot X^{\gamma}$$

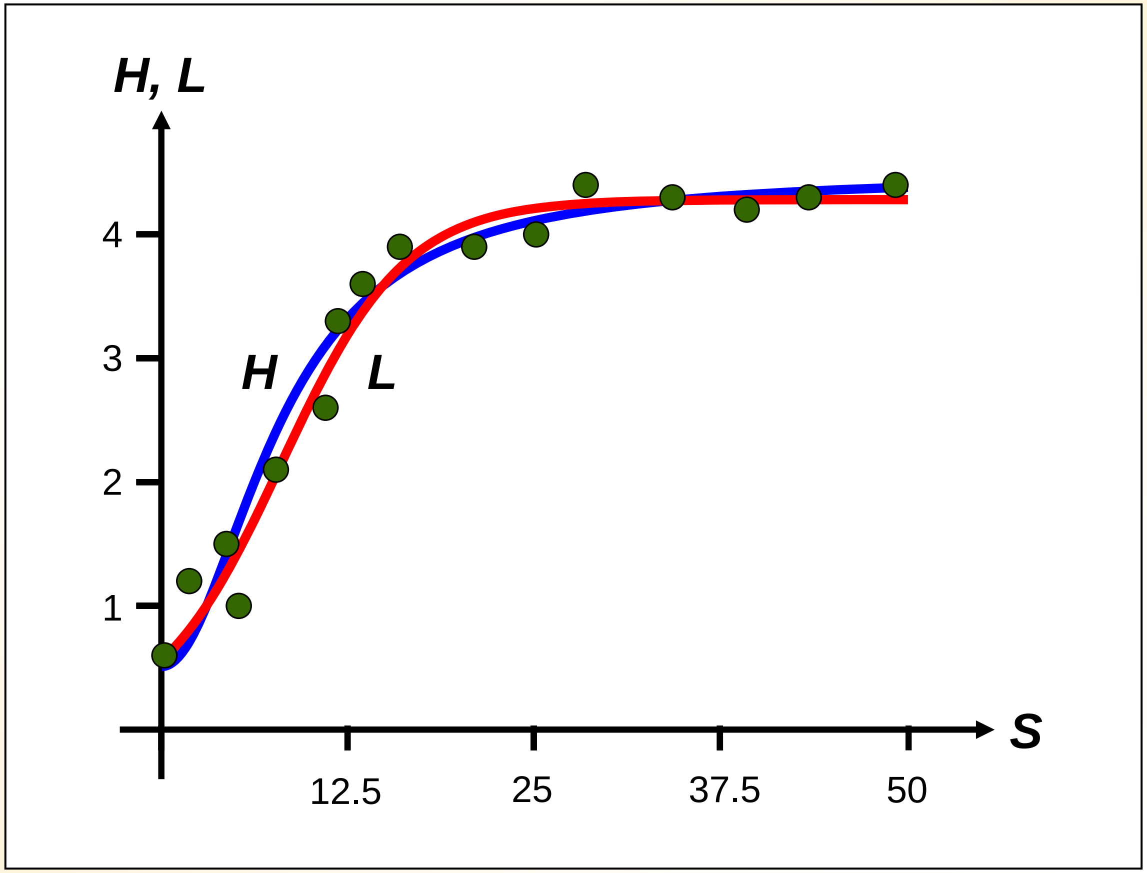
α =1.75 and γ = 0.8

$DF/dt = \ldots f(X,Y)\ldots$

Example:

$$\begin{aligned} f &= 2.45 \cdot X^{1.2} \cdot Y^{-0.3} \\ &= 2.45 \cdot X^{1.2} \cdot Y \cdot Y^{-1.3} \\ &= 2.45 \cdot X^{1.2} \cdot (1.75 \cdot X^{0.8}) \cdot Y^{-1.3} \\ &= 4.2875 \cdot X^{2} \cdot Y^{-1.3} \end{aligned}$$

Similar Fits with Different Models

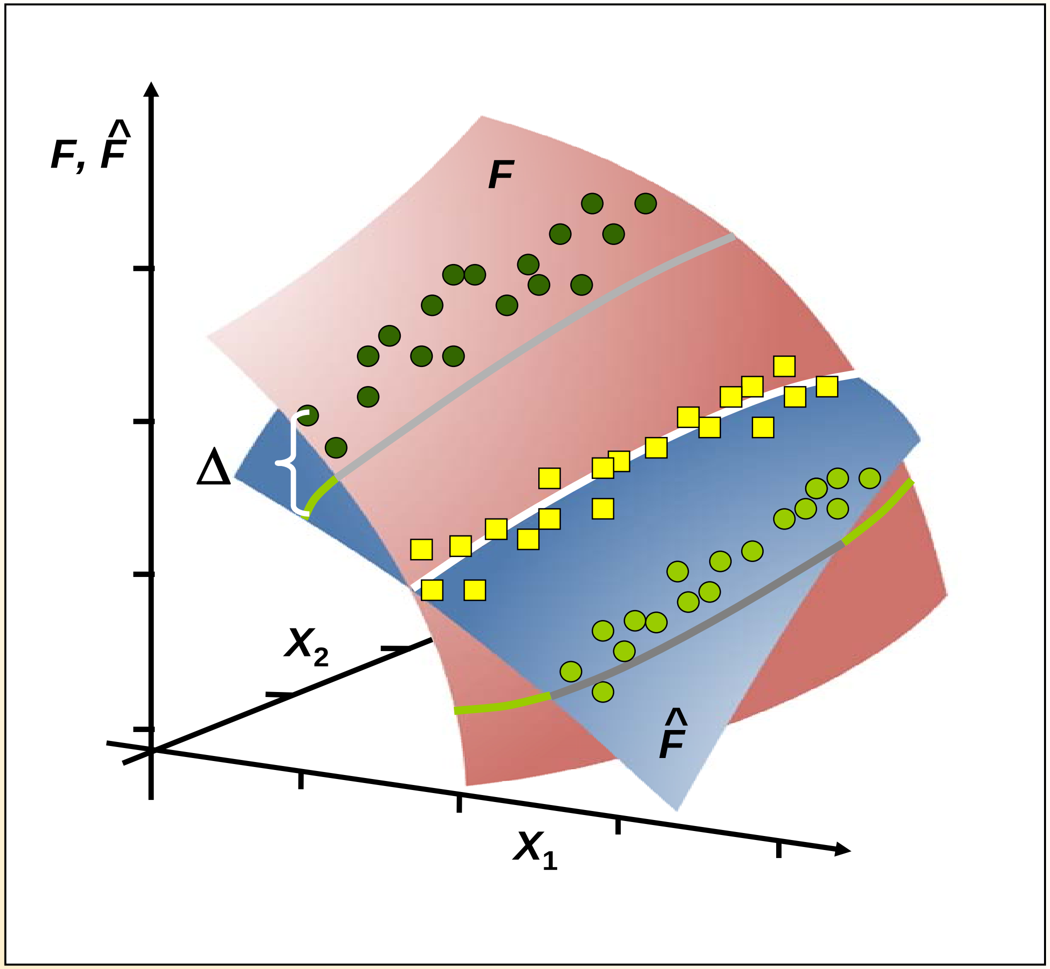


$$H(S) = 4S^2/(8^2 + S^2) + 0.5$$

$$L(S) = 4.3/[1 + \exp(-0.24 \cdot (S - 8))]$$

Voit, 2010

Insufficiently Informative Data



One data set (yellow)

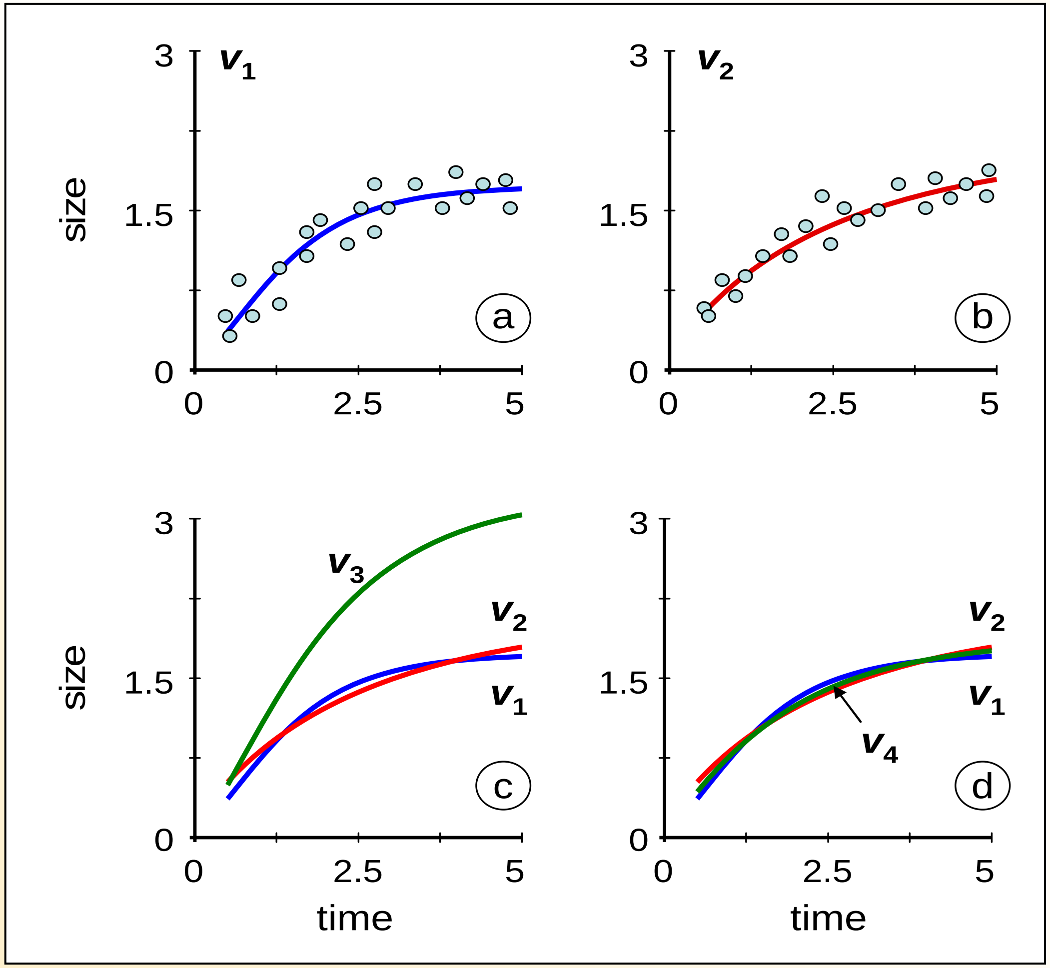
Fit yellow data with function $F(X_1, X_2)$:
White line in 3-dim space
Line is part of red surface

True model: Blue surface $\hat{F}(X_1, X_2)$

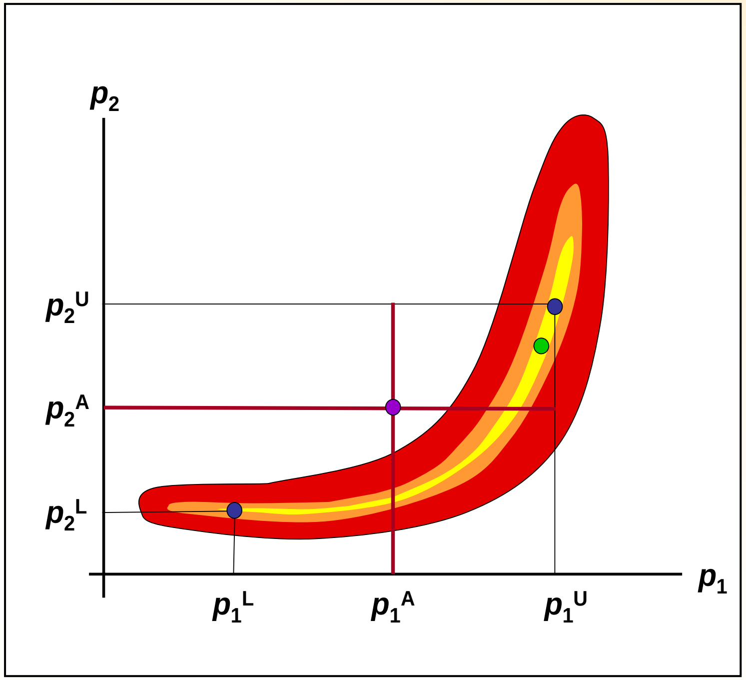
Extrapolation with $F(X_1, X_2)$ bad for green data

Voit, 2010

Averaging of Estimation Results

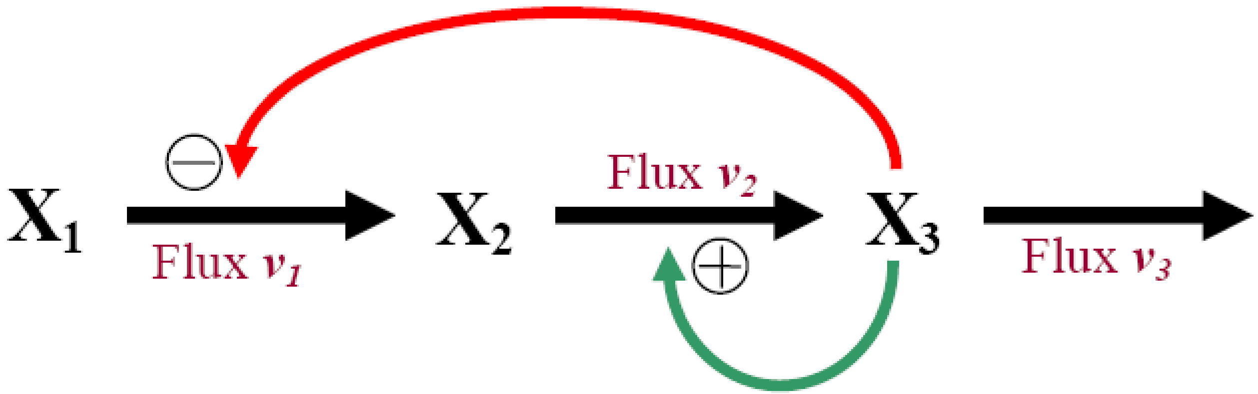


$$w(t) = (p_1 - p_2 \cdot \exp(-p_3 \cdot t))^{p_4}$$



Voit, 2012

Problems with Compensation



$$X_1 = \text{Constant}$$

$$\dot{X}_2 = \frac{(X_1) * V_{\max}}{K_m\left[1 + \frac{X_3}{K_i}\right] + X_1} - p_1 X_2^{\,p_2} X_3^{p_3}$$

$$\dot{X}_3 = p_1 X_2^{\,p_2} X_3^{p_3} - p_4 X_3^{p_5}$$

Goel, Chou, Voit, *Bioinformatics*, 2008

Problems with Compensation

Table S1: Error compensation within the same flux (v_1)

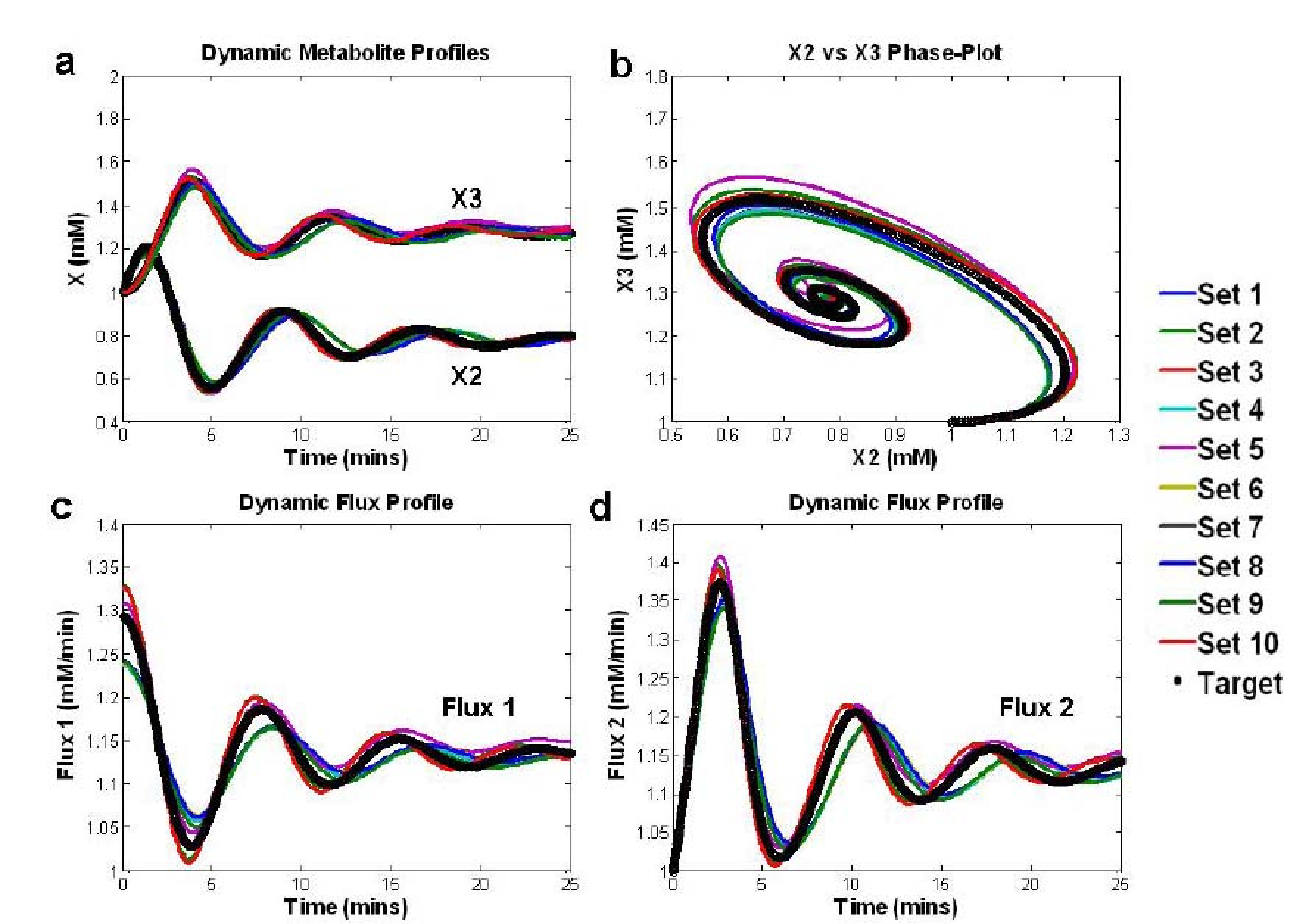
Set	V_{max}	K_m	K_i	p_1	p_2	p_3	p_4	p_5	Residual
1	88.2533	91.2397	1.8482	1	0.5	1	1	0.5	6.3238
2	18.6819	9.7831	0.5992	1	0.5	1	1	0.5	2.0628
3	63.0698	66.1785	1.9714	1	0.5	1	1	0.5	7.0341
4	91.0532	94.3597	1.855	1	0.5	1	1	0.5	6.4499
5	14.2804	10	1.019	1	0.5	1	1	0.5	3.8237
6	82.7704	87.9852	2.0162	1	0.5	1	1	0.5	7.3094
7	88.7362	93.0726	1.9447	1	0.5	1	1	0.5	6.6048
8	92.4504	97.0702	1.9466	1	0.5	1	1	0.5	6.616
9	68.9295	67.7172	1.6343	1	0.5	1	1	0.5	4.9066
10	18.2178	8.9871	0.5458	1	0.5	1	1	0.5	2.2876

$$X_1 = \text{Constant}$$

$$\dot{X}_2 = \frac{(X_1) * V_{\max}}{K_m\left[1 + \frac{X_3}{K_i}\right] + X_1} - p_1 X_2^{p_2} X_3^{p_3}$$

$$\dot{X}_3 = p_1 X_2^{p_2} X_3^{p_3} - p_4 X_3^{p_5}$$

Problems with Compensation



Problems with Compensation

Table S2: Error compensation between fluxes (v_1 and v_2)

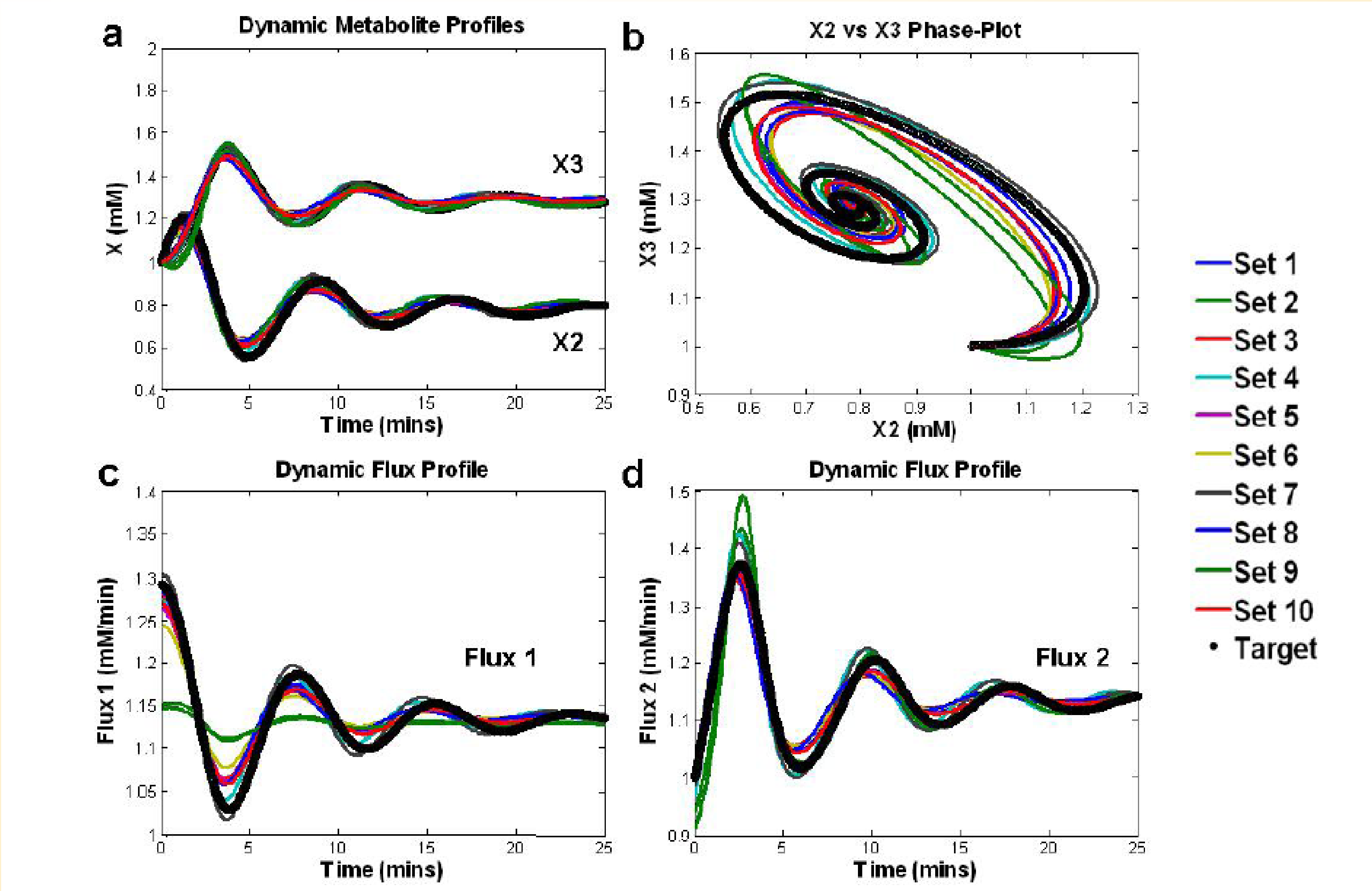
Set	V_{max}	K_m	K_i	p_1	p_2	p_3	p_4	p_5	Residual
1	104.9701	92.1829	1.3281	1.0021	0.5785	1.0038	1	0.5	3.4688
2	57.0719	91.5615	15.2508	0.9401	0.9865	1.7386	1	0.5	4.4663
3	13.0088	9.5706	1.0968	1.0173	0.5921	0.9671	1	0.5	6.6559
4	103.6876	93.837	1.3967	0.9688	0.6418	1.2038	1	0.5	5.6134
5	12.4525	9.971	1.2927	1.0055	0.5812	1.0271	1	0.5	2.8754
6	10.01	8.8733	1.7075	1	0.6676	1.1052	1	0.5	6.624
7	124.476	88.9055	0.8893	0.9841	0.544	1.0853	1	0.5	3.0074
8	13.5262	9.5896	1.0152	1.013	0.6045	1.0017	1	0.5	7.2336
9	60.7643	96.3775	13.346	0.9117	1.0602	1.8375	1	0.5	6.3344
10	12.3914	9.5007	1.1869	1.0086	0.5676	1.0079	1	0.5	2.7299

$$X_1 = \text{Constant}$$

$$\dot{X}_2 = \frac{(X_1) * V_{\max}}{K_m\left[1 + \frac{X_3}{K_i}\right] + X_1} - p_1 X_2^{p_2} X_3^{p_3}$$

$$\dot{X}_3 = p_1 X_2^{p_2} X_3^{p_3} - p_4 X_3^{p_5}$$

Problems with Compensation



Problems with Compensation

Table S3: Error compensation among different equations (v_1 and v_3)

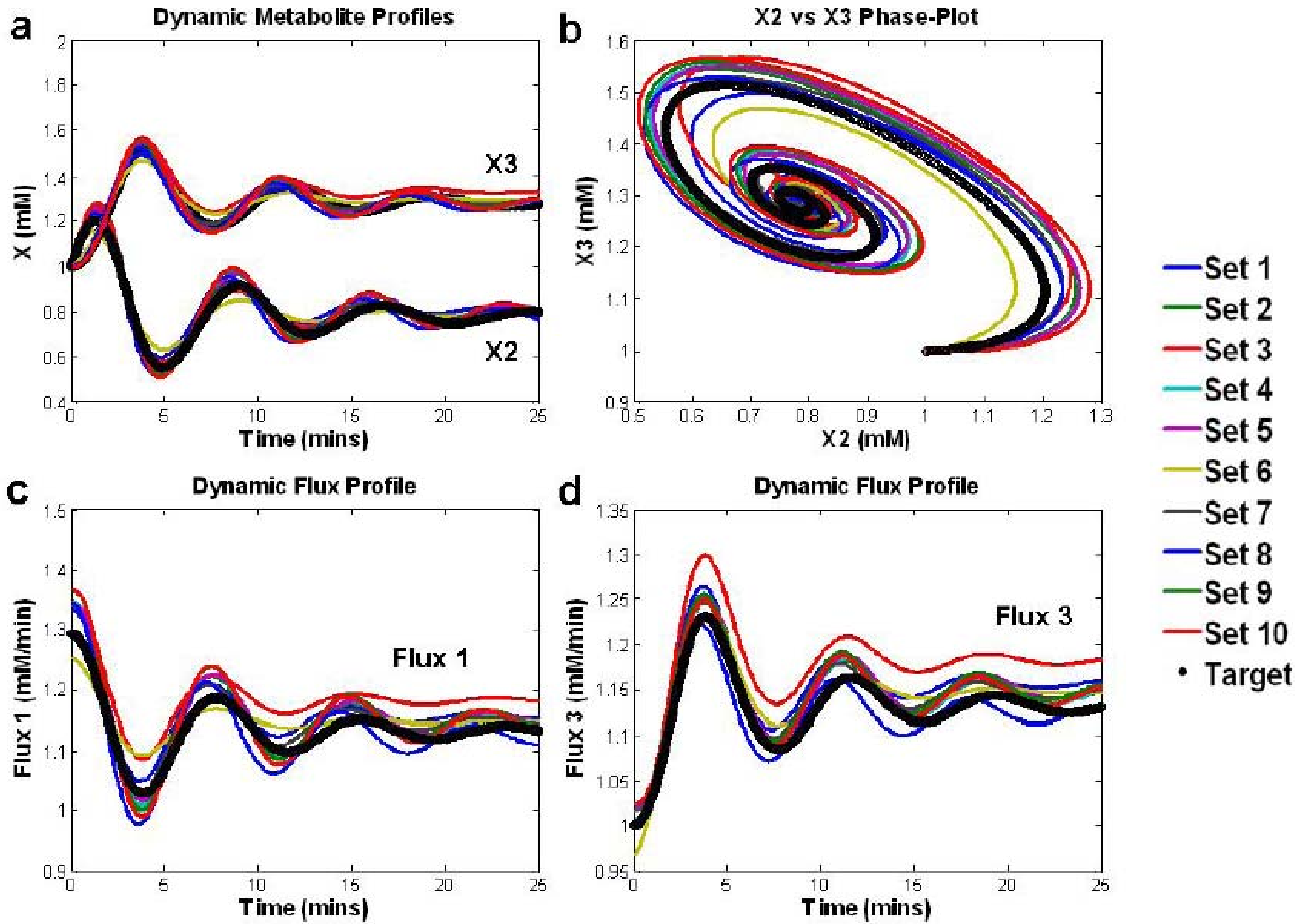
Set	V_{max}	K_m	K_i	p_1	p_2	p_3	p_4	p_5	Residual
1	17.5775	9.9988	0.6979	1	0.5	1	1.001	0.5786	4.5287
2	19.0012	9.0003	0.5203	1	0.5	1	1.0178	0.4659	3.2879
3	11.0985	7.5279	1.0842	1	0.5	1	1.0001	0.5842	7.1035
4	16.5287	7.7719	0.5241	1	0.5	1	1.0205	0.4605	3.5256
5	17.8896	9.2186	0.5967	1	0.5	1	1.0206	0.4705	3.1041
6	87.5991	94.1804	2.1613	1	0.5	1	0.9669	0.6658	5.1819
7	15.5174	7.7989	0.5839	1	0.5	1	1.0011	0.5316	2.5845
8	24.2938	8.3902	0.3257	1	0.5	1	1.0057	0.4595	7.4577
9	21.3578	9.055	0.4464	1	0.5	1	1.0248	0.4567	6.3633
10	22.064	8.7065	0.4023	1	0.5	1	1.0256	0.4397	7.1653

$$X_1 = \text{Constant}$$

$$\dot{X}_2 = \frac{(X_1) * V_{max}}{K_m\left[1 + \frac{X_3}{K_i}\right] + X_1} - p_1 X_2^{p_2} X_3^{p_3}$$

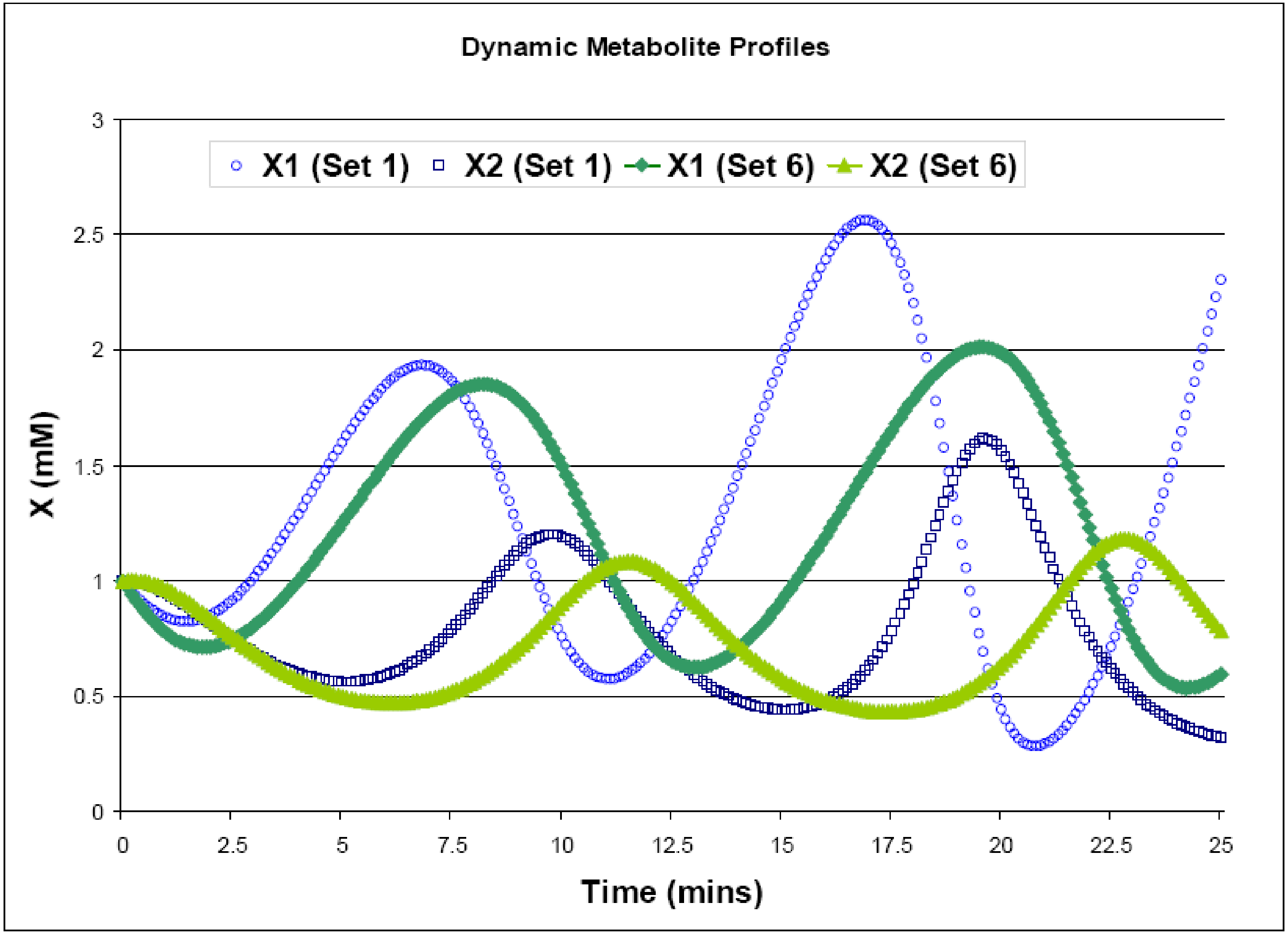
$$\dot{X}_3 = p_1 X_2^{p_2} X_3^{p_3} - p_4 X_3^{p_5}$$

Problems with Compensation



Problems with Compensation

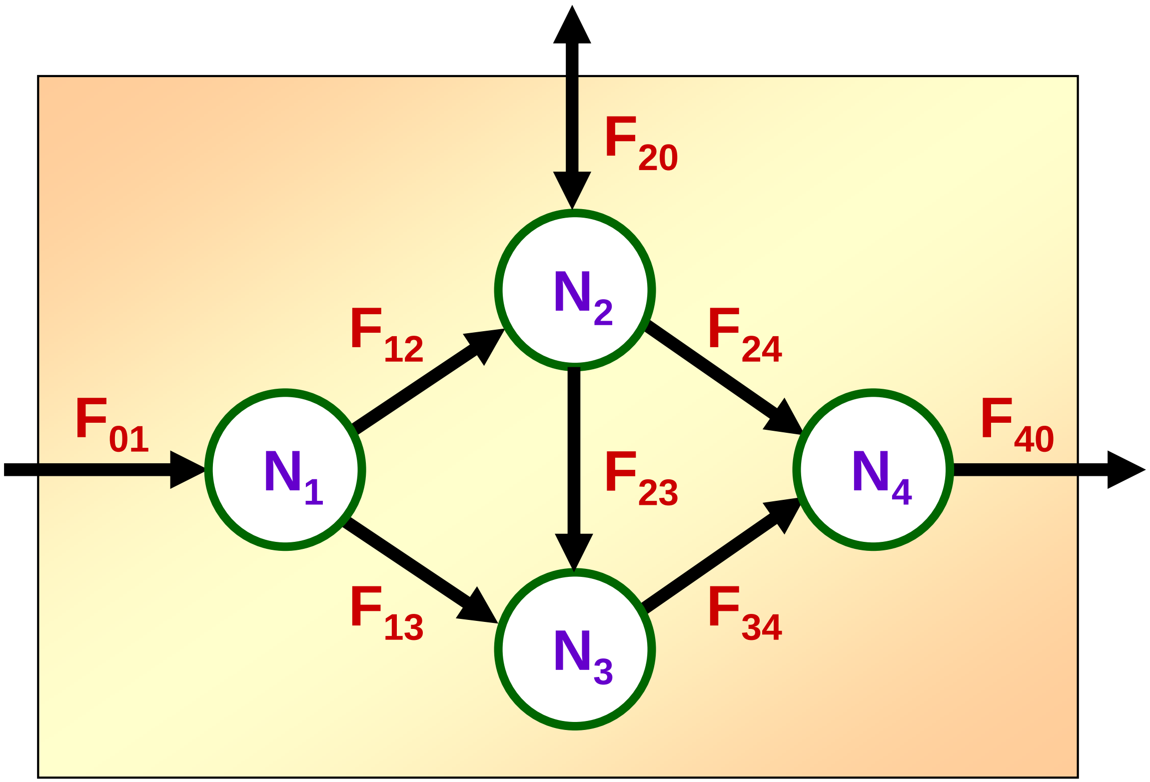
Mild extrapolation: Reduce input X_1 from 2 to 1.1



Dynamic Flux Estimation (DFE)

Inspired by Stoichiometric and Flux Balance Analysis (purely at steady state)

Extended to dynamic time courses: $\frac{dX_i}{dt} = \dot{X}_i = \sum Influxes - \sum Effluxes.$



$$\frac{dN_1}{dt} = F_{01} - F_{12} - F_{13}$$

$$\frac{dN_2}{dt} = F_{12} + F_{20F} - F_{20R} - F_{23} - F_{24}$$

$$\frac{dN_3}{dt} = F_{13} + F_{23} - F_{34}$$

$$\frac{dN_4}{dt} = F_{24} + F_{34} - F_{40}$$

Dynamic Flux Estimation (DFE)

Concept:

Study flux balance at each time point

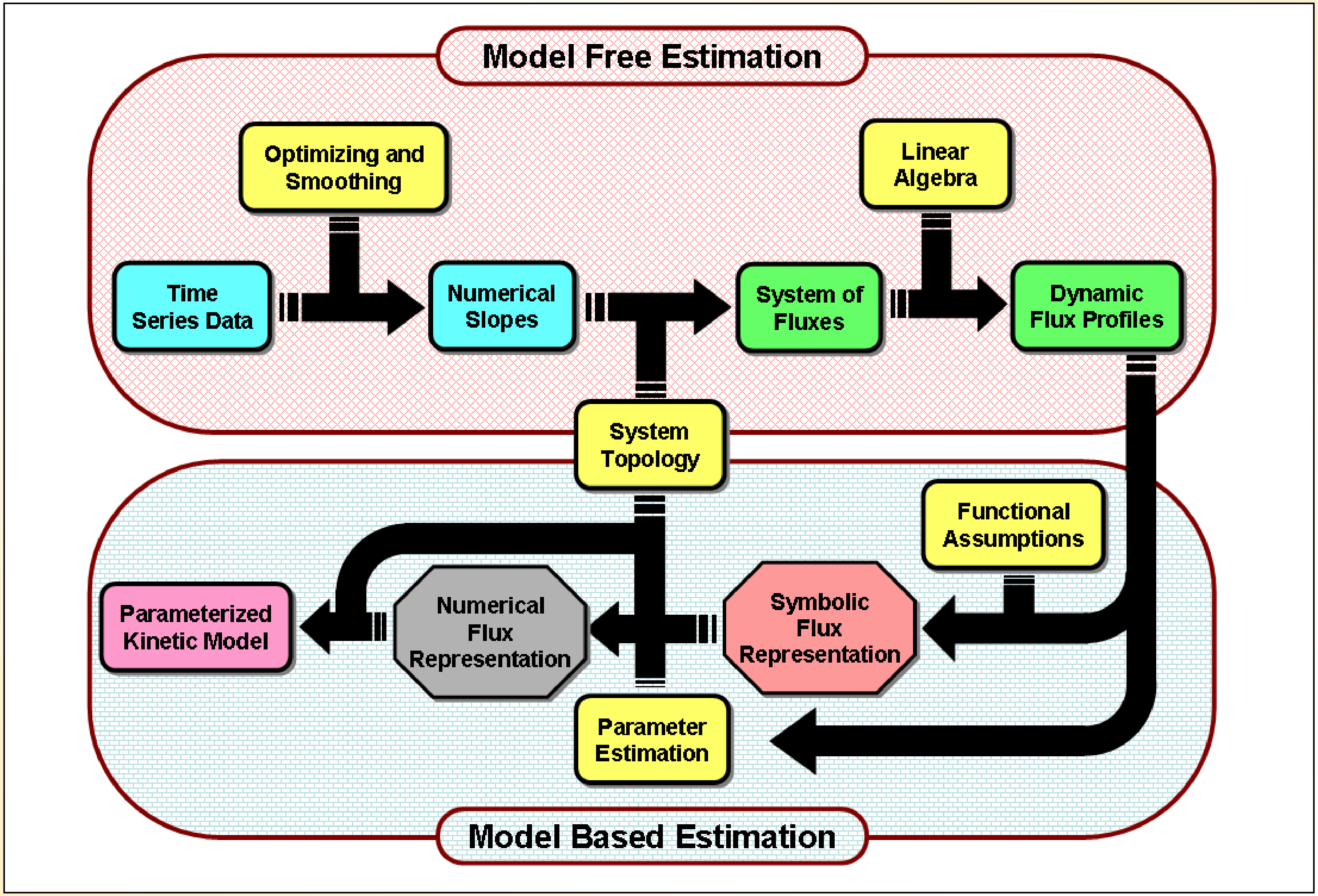
Change in variable @ t = All influxes @ t – All effluxes @ t

Linear system; solve as far as possible

Result: values of each flux @ time points t_i (non-parametric; no functional forms!)

Represent fluxes with appropriate models

Dynamic Flux Estimation (DFE)



Goel, Chou, Voit, *Bioinformatics*, 2008

Problems with DFE

Issue 1: The connectivity (reactions and/or regulation) of the system is not fully known.

Issue 2: Some time series were not measured, although metabolites are involved in the pathway.

Issue 3: Some unknown or not measured metabolites are important.

Issue 4: The flux system is under-determined. This situation is the rule rather than the exception.

Voit et al., IET Systems, 2010

Solution Strategies

Issue 1: ***The connectivity (reactions and/or regulation) of the system is not fully known.***

Causality models

Correlation-based approaches

Fitting alternative candidate models

Fitting superstructures (families of models that contain special cases)

Biochemical Systems Theory or other canonical models useful

Requires very good data

Solution Strategies

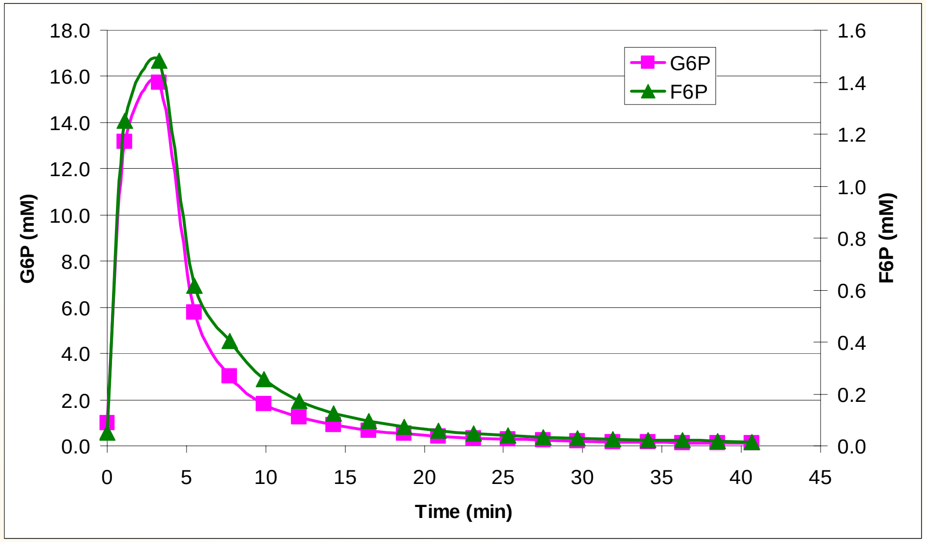
Issue 2: *Some time series were not measured, although metabolites are involved in the pathway*

Mass negligible?

Information about reactions associated with missing metabolite?

Example: reversible isomerization of G6P (measured) to F6P (not measured)

$$v_2 = \frac{v_{\max}^{for} \cdot \frac{[G6P]}{K_{mG6P}} - v_{\max}^{rev} \cdot \frac{[F6P]}{K_{mF6P}}}{1 + \frac{[G6P]}{K_{mG6P}} + \frac{[F6P]}{K_{mF6P}} + \frac{[P_i]}{K_{mP_i}}}$$



***In vivo* NMR measurements of G6P in *Lactococcus lactis* (literature) and time series of F6P (scaled) reconstructed with kinetic literature information**

Solution Strategies

Issue 3: *Some unknown or not measured metabolites are important*

Affecting pertinent mass? (C versus P or H; G6P ~ F6P; NAD^+ ~ NADH)

Mass balanced? (Total mass over time ~ constant?)

Yes: metabolites may be ignorable

No: problem with no good solution

Solution Strategies

Issue 4: ***The flux system is under-determined. This situation is the rule rather than the exception***

Determine some fluxes with other means

Kinetic information

New method:

Estimate enough fluxes from time series data

to render the system full rank

Chou and Voit, *BMC Syst. Biol.,* 2012

Individual Flux Estimation

Basic Concept: Consider simple dynamics of X_i

$$X_j \rightarrow X_i \rightarrow$$

$$\dot{X}_i = v_i^+(X_j) - v_i^-(X_i)$$

Assume that v_i^- is a function in a strict mathematical sense.
Look for time points (in the same or in similar datasets) where X_i has the same value (*e.g.*, c_i), whereas X_j has a different value at each of these time points. If so, all values of v_i^- are the same: vc_i

$$\dot{X}_i = v_i^+(X_j) - vc_i$$

Observe $\dot{X}_i$ at several time points; point-estimate v_i^+

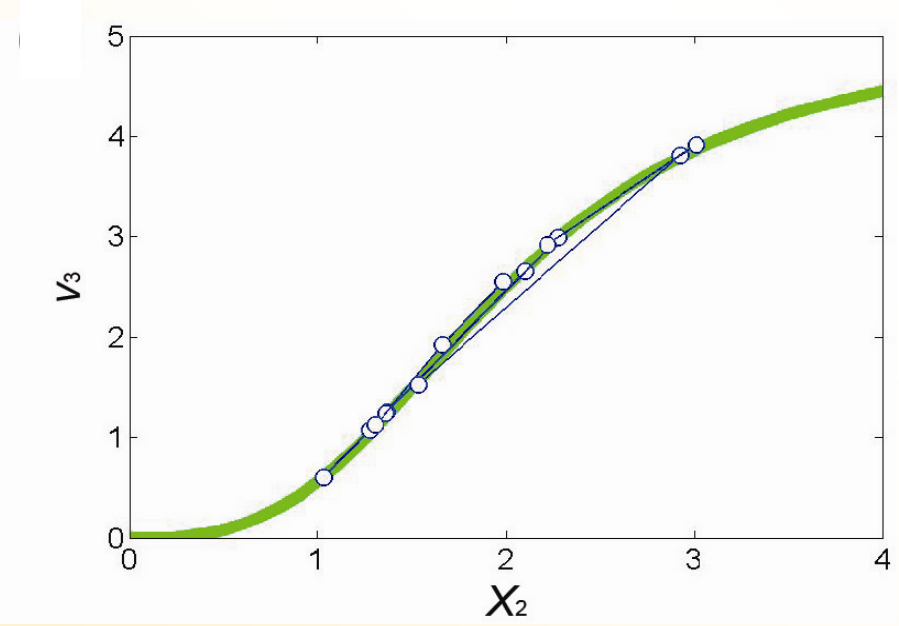
Individual Flux Estimation

Result: point-estimates of v_i^+

Can plot these estimates against time or against dependent variable

No functional form!

Functional form may be estimated in second step



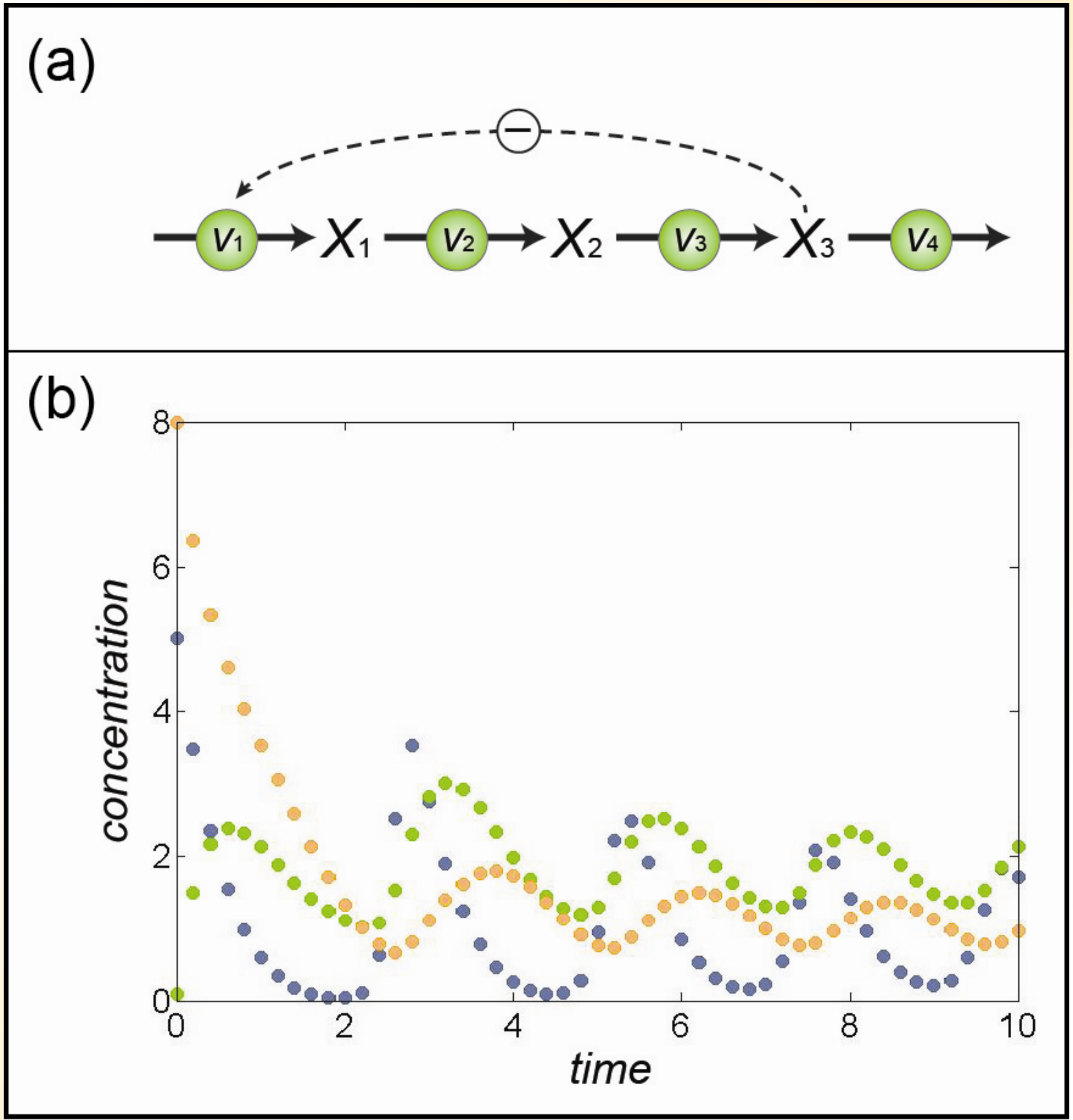
Individual Flux Estimation

Example

$$\dot{X}_1 = v_1 - v_2$$
$$\dot{X}_2 = v_2 - v_3.$$
$$\dot{X}_3 = v_3 - v_4$$

Unknown fluxes

$$v_1 = 1.5\, X_3^{-6}$$
$$v_2 = 2.4\, X_1^{0.8}$$
$$v_3 = \frac{V_{\max} X_2^3}{K_M^3 + X_2^3},$$
$$v_4 = 2\, X_3^{0.75}$$



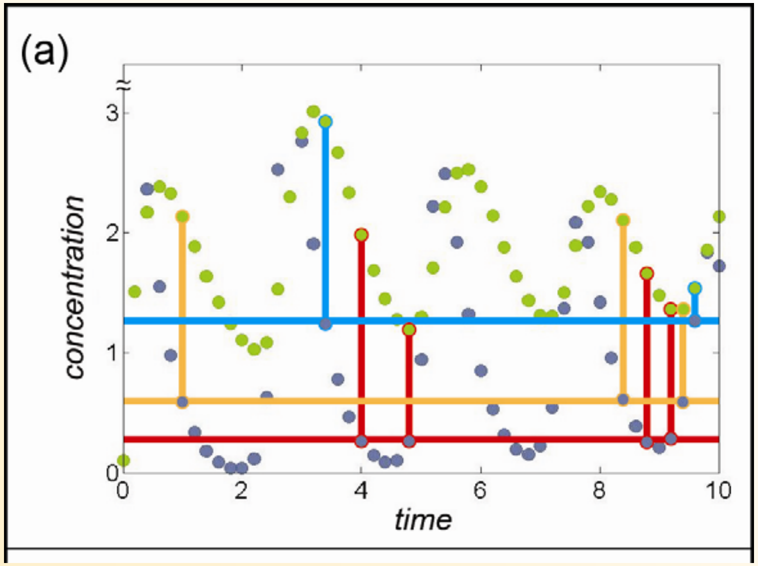
Individual Flux Estimation

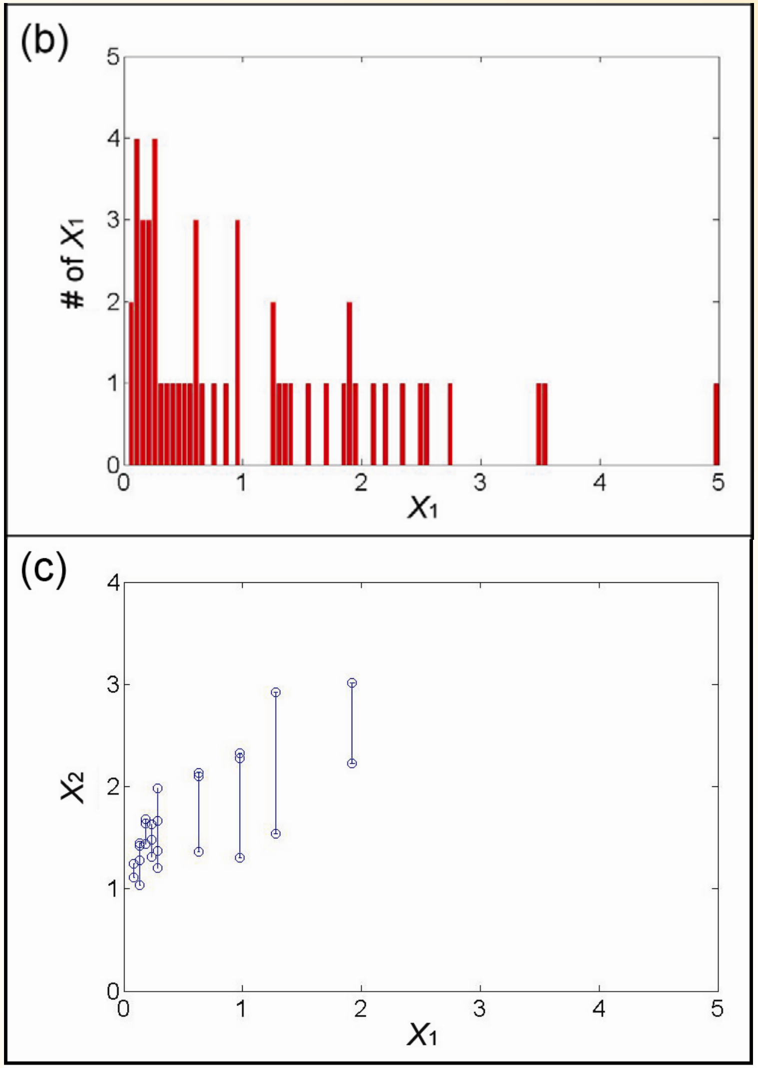
Collect data where X_1 has the same value

Bin values

Assign X_2 values to binned X_1 values

Estimate slopes S_2 (= derivatives of X_2)





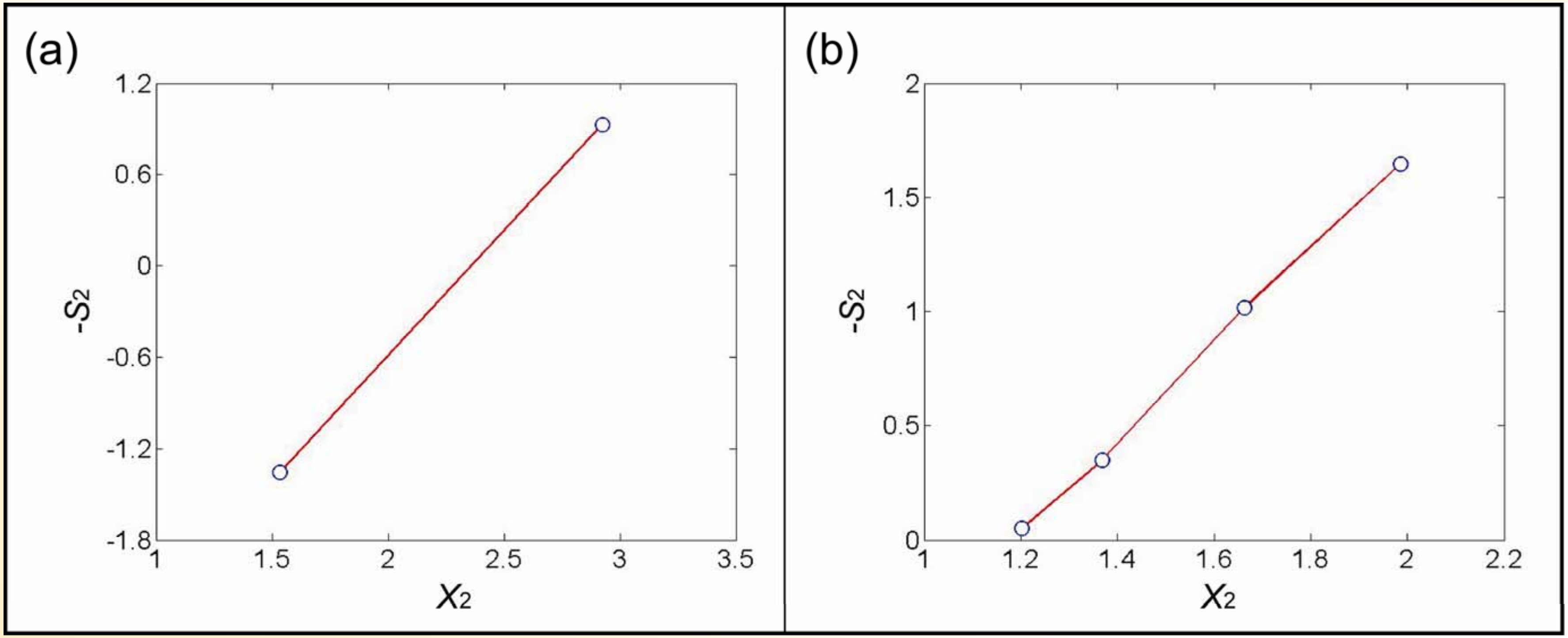
Individual Flux Estimation

Recall equation of X_2

$$\dot{X}_2 = v_2 - v_3.$$

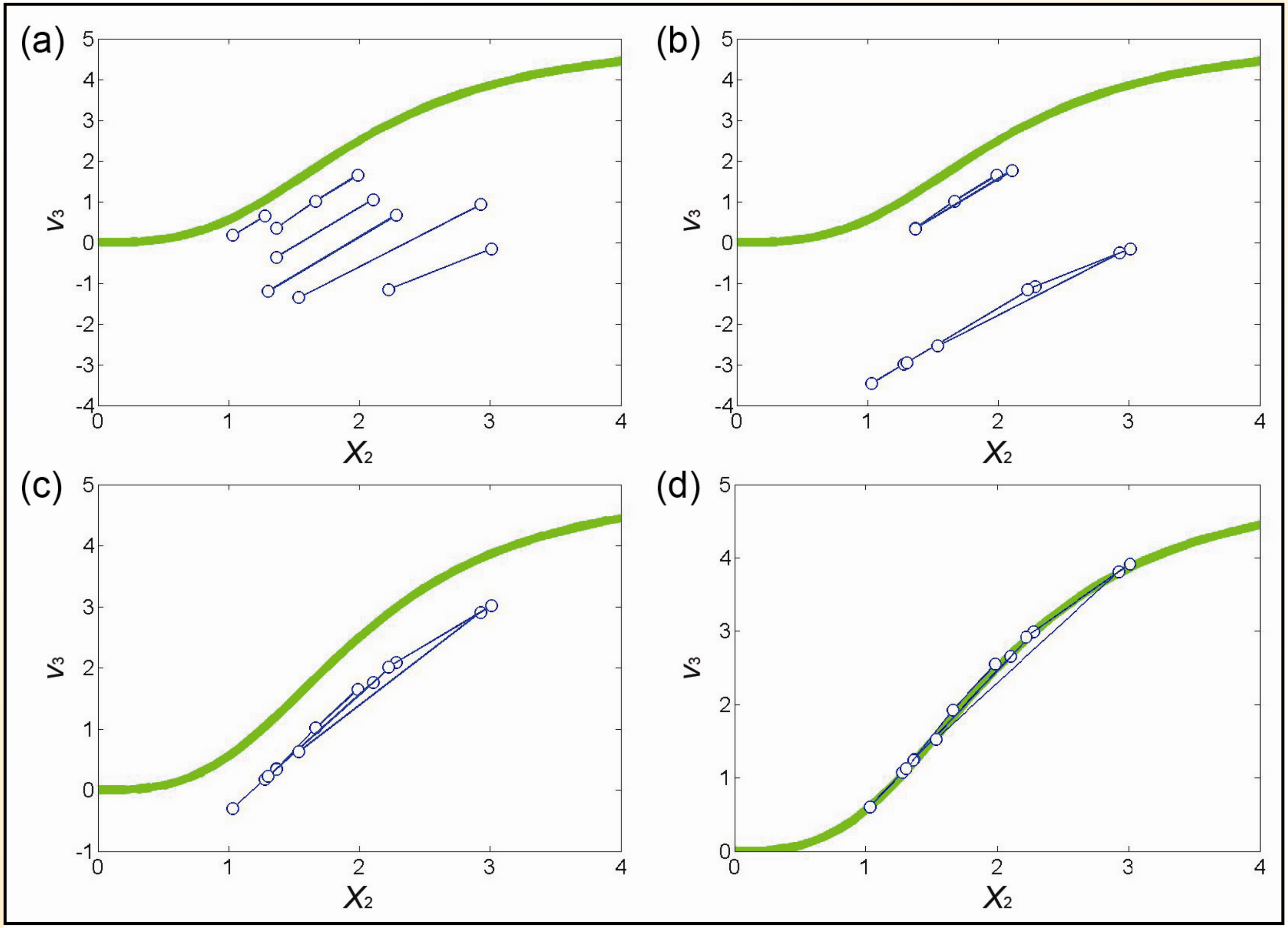
For X_1 with equal value, $v_2 = 2.4\, X_1^{0.8}$ must have the same (but unknown) value

Estimate slopes S_2 from data; point-estimate v_3



Individual Flux Estimation

Repeat for many sets of X_1 values; shift as needed; e.g., $v(0) = 0$

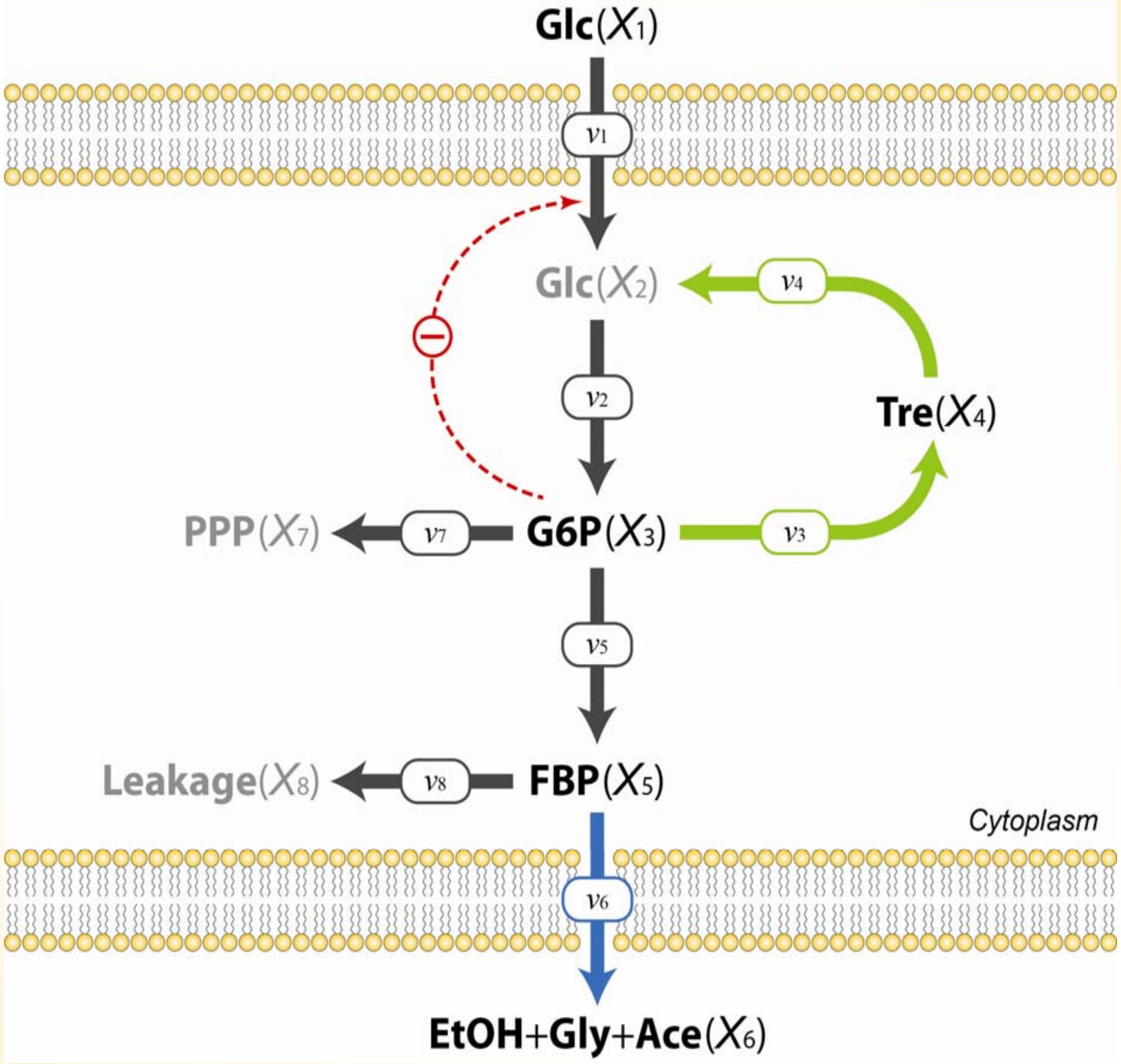


Example: Trehalose Pathway

Flux system (functions unknown)

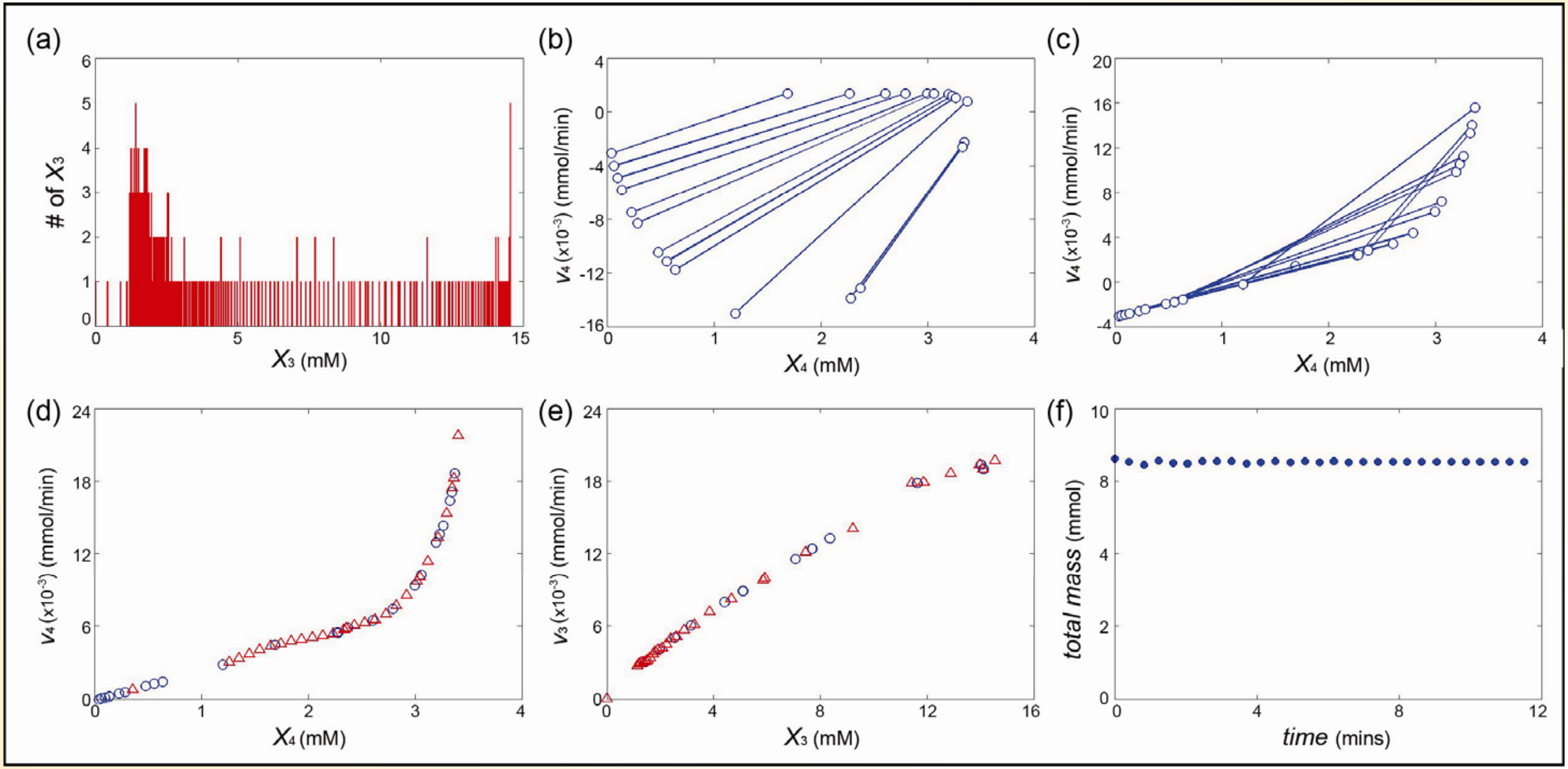
$$\dot{X}_1 = -v_1 / V_{ext}$$
$$\dot{X}_2 = (v_1 + 2v_4 - v_2) / V_{int}$$
$$\dot{X}_3 = (v_2 - 2v_3 - v_5 - v_7) / V_{int}$$
$$\dot{X}_4 = (v_3 - v_4) / V_{int}$$
$$\dot{X}_5 = (v_5 - v_6 - v_8) / V_{int}$$
$$\dot{X}_6 = 2v_6 / V_{ext}$$
$$\dot{X}_7 = v_7 / V_{int}$$
$$\dot{X}_8 = v_8 / V_{int}$$

Rank deficiency = 1



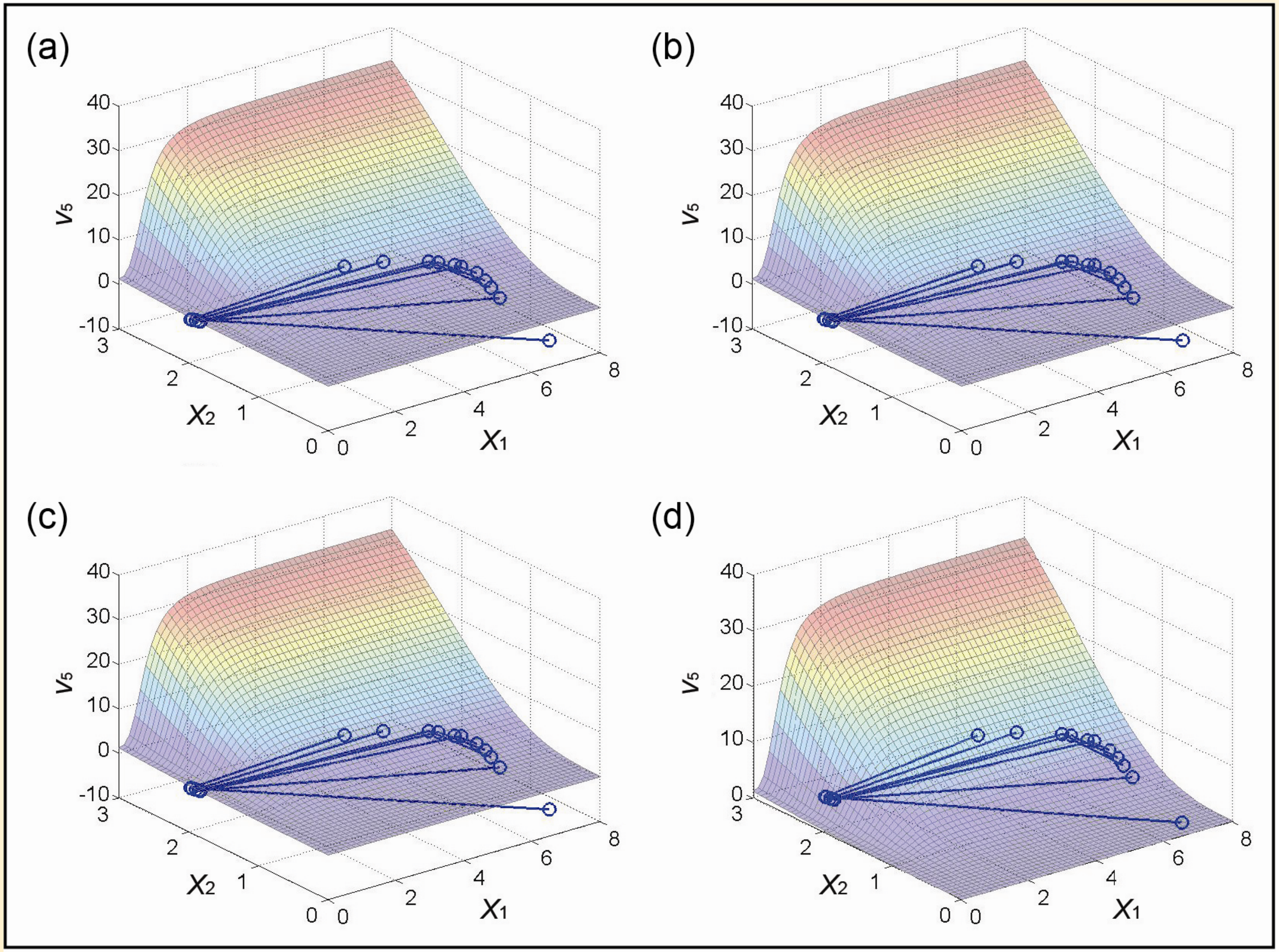
Example: Trehalose Pathway

Same procedure as before for one flux of our choice; here v_4
Once one flux estimated, system has full rank



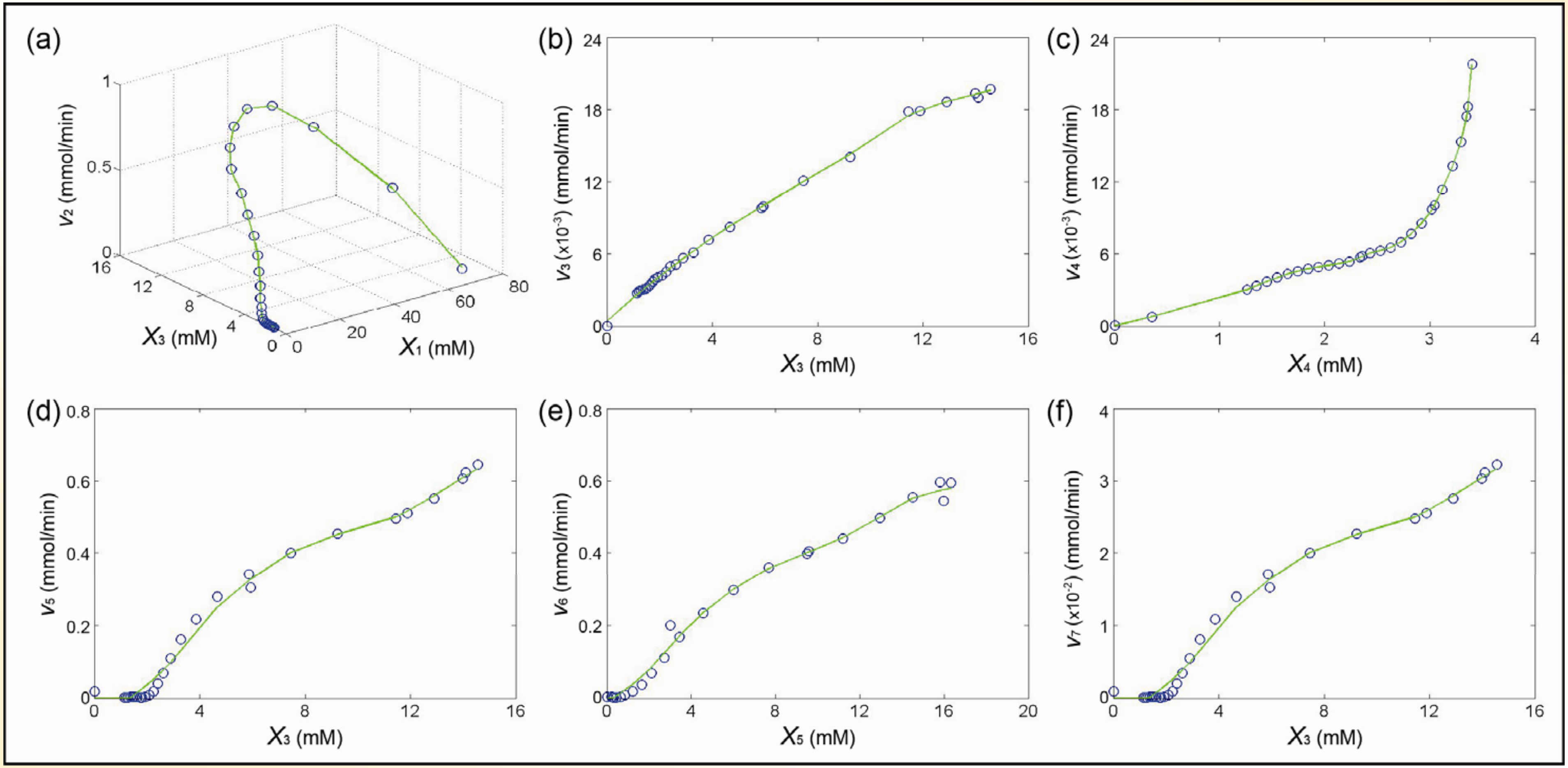
Individual Flux Estimation

Same principles for fluxes depending on two metabolites



Example: Trehalose Pathway

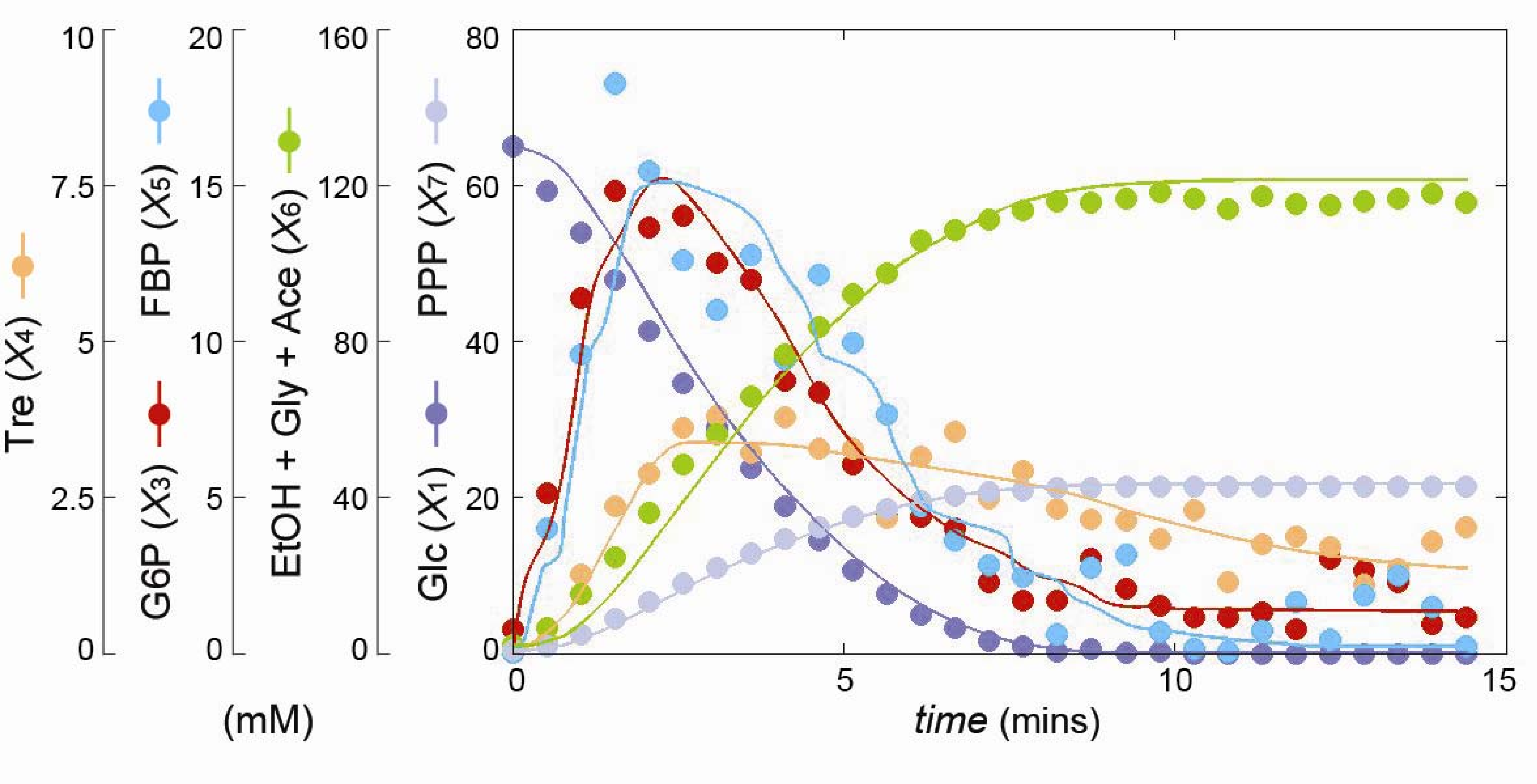
Estimated fluxes



Shapes (vs. time and vs. metabolites) characterized
Functional representations unknown (non-parametric estimation)

Example: Trehalose Pathway

Reconstruction of dynamics, using estimated fluxes (functional forms unknown)



Summary and Acknowledgments

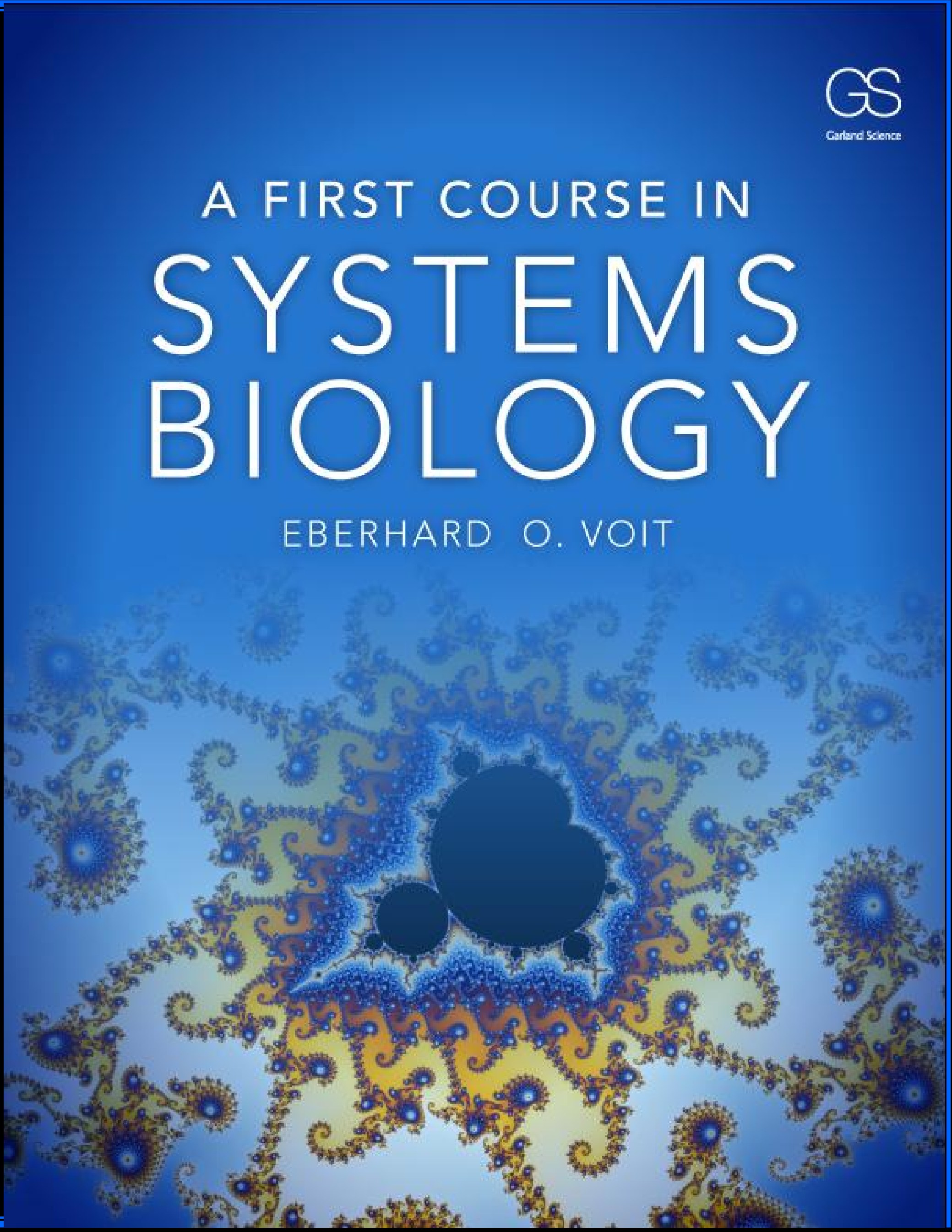
- Parameter estimation complicated (bottleneck of modeling)
- Quality of fit (defined as residual error) not sufficient
- Parameter estimation even more complicated if functions unknown
- DFE works well, if enough data are available and system full rank
- If not, parametric tricks
- Filling rank possible if suitable data available

I-Chun Chou

Funding: NIH, NSF, DOE, Woodruff Foundation,
University System of Georgia,
Georgia Research Alliance

Information: www.bst.bme.gatech.edu

Just published!



References

Schulz, A.R.:Enzyme Kinetics: From Diastase to Multi-enzyme Systems, Cambridge University Press, Cambridge; New York, 1994.

Voit, E.O.: *Computational Analysis of Biochemical Systems. A Practical Guide for Biochemists and Molecular Biologists*, Cambridge University Press, Cambridge, U.K., 2000.

Voit, E.O.: The dawn of a new era of metabolic systems analysis, *Drug Discovery Today BioSilico* **2(5)**, 182-189, 2004.

Voit, E.O., J.S. Almeida, S. Marino, R. Lall, G. Goel, A.R. Neves, and H. Santos. Regulation of glycolysis in *Lactococcus lactis*: An unfinished systems biological case study. *IEE Proc. Systems Biol.* **153**, 286-298, 2006.

Goel, G., I-C. Chou, and E.O. Voit: System estimation from metabolic time series data. *Bioinformatics* **24**, 2505-2511, 2008

Vilela, M., S. Vinga, M. A. Grivet Mattoso Maia, E. O. Voit, J. S. Almeida: Identification of neutral sets of biochemical systems models from time series data. BMC Systems Biology **3**:47, 2009.

Voit, E.O., G. Goel , I-C. Chou, and L. da Fonseca: Estimation of metabolic pathway systems from different data sources. *IET Systems Biol.* **3(6)**, 513-522, 2009.

Voit, E.O. and I-C. Chou: Parameter estimation in canonical biological systems models. *Int. J. Syst. Synth. Biol.* **1(1)**, 1-19, 2010.

Voit, E.O.: What if the fit is unfit? Criteria for biological systems estimation beyond residual errors. In: M. Dehmer, F. Emmert-Streib and A. Salvador (Eds.): Applied Statistics for Biological Networks. J. Wiley and Sons, New York, pp. 183-200, 2011.

Chou, I-C., and E.O. Voit: Estimation of dynamic flux profiles from metabolic time series data. *BMC Systems Biology*, in press 2012.

Voit, E.O.: *A First Course in Systems Biology*. Garland Science, New York, NY, 2012.