Parameter Identifiability of Metabolic Network Models

26/07/2010

Sridharan Srinath & Rudiyanto Gunawan

Department of Chemical and Biomolecular Engineering
National University of Singapore
Dynamical Bio Systems Laboratory

*srinath@nus.edu.sg

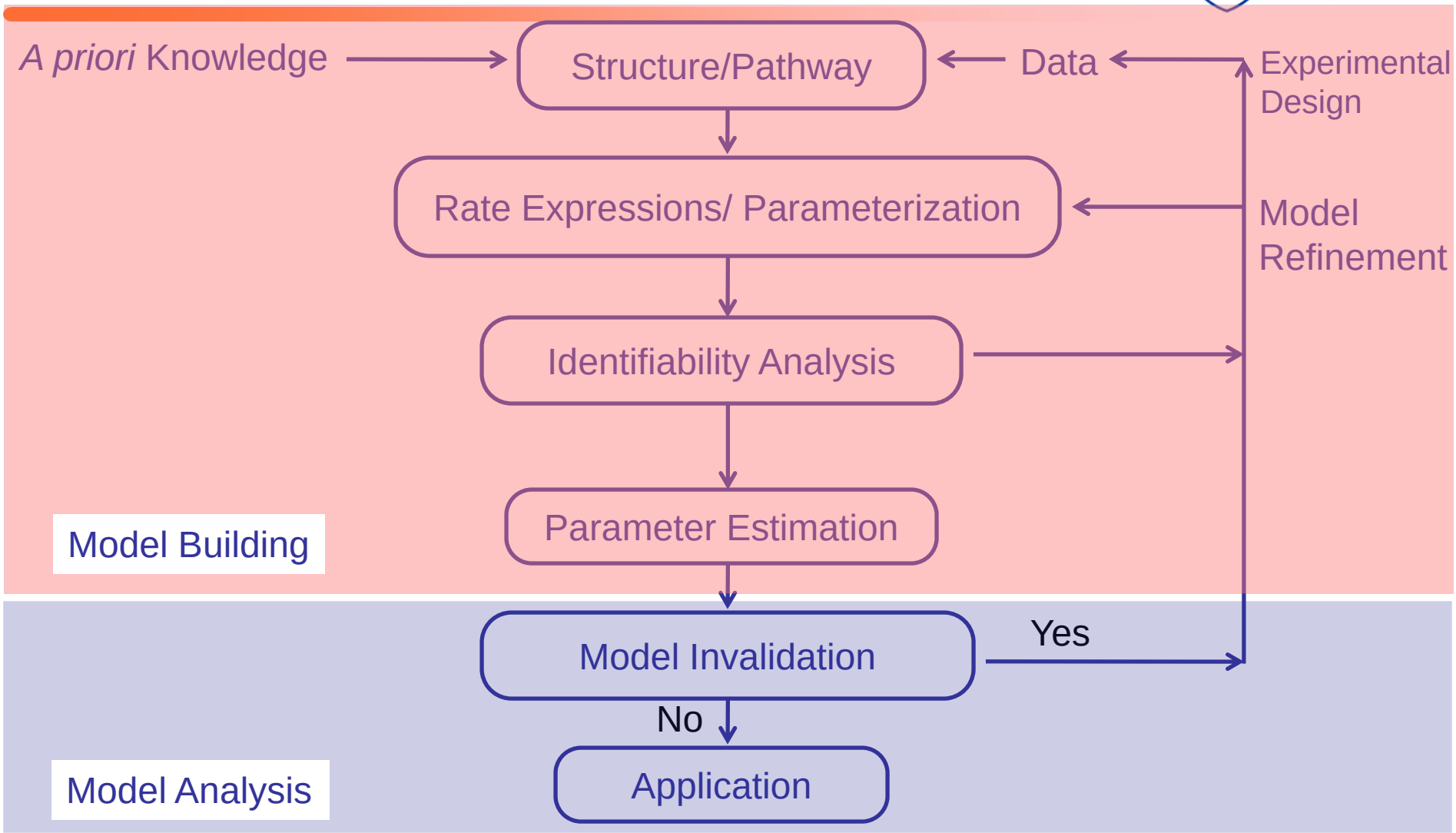
Objectives

- To develop method(s), based on statistics, to perform identifiability analysis (IA) of dynamic models.

- Apply IA on few metabolic network models

  (*will talk about the case studies only if time permits; will mainly focus on the method*)

Model Identification Cycle



A priori or Structural IA

- Given perfect data, is it possible to uniquely determine the values of the model parameters?
- Consider a simple system $y = (\theta_1 + \theta_2) x_1$

  the parameters θ_1 and θ_2 are NOT independently *a priori* identifiable
- In another system: $y = \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2$
- Given measurements of (x_1, x_2, y)
  - from experiments with $x_1 = 0$, θ_2 is *a priori* identifiable
  - from experiments with $x_2 = 0$, θ_1 is *a priori* identifiable
  - Generally, parameters are *a priori* identifiable if model is "invertible"

$$\mathbf{y} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_1\mathbf{x}_2 \end{bmatrix} \boldsymbol{\theta} = \mathbf{X}\boldsymbol{\theta} \quad \Rightarrow \quad \boldsymbol{\theta} = \tilde{\mathbf{X}}^{-1}\mathbf{y}$$

  - For inverse to exist, $\begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_1\mathbf{x}_2 \end{bmatrix}$ should have full column rank

Sensitivity analysis

$$\frac{\partial \mathbf{y}(t_n, x_n, \boldsymbol{\theta})}{\partial \theta_p}; \quad p=1,2,...P$$

$$\mathbf{S} = \begin{pmatrix} \left.\frac{\partial \mathbf{y}_1}{\partial \theta_1}\right|_{t=t_1} & \cdots & \left.\frac{\partial \mathbf{y}_1}{\partial \theta_P}\right|_{t=t_1} \\ \vdots & \ddots & \vdots \\ \left.\frac{\partial \mathbf{y}_R}{\partial \theta_1}\right|_{t=t_1} & \cdots & \left.\frac{\partial \mathbf{y}_R}{\partial \theta_P}\right|_{t=t_1} \\ \left.\frac{\partial \mathbf{y}_1}{\partial \theta_1}\right|_{t=t_2} & \cdots & \left.\frac{\partial \mathbf{y}_1}{\partial \theta_P}\right|_{t=t_2} \\ \vdots & \ddots & \vdots \\ \left.\frac{\partial \mathbf{y}_R}{\partial \theta_1}\right|_{t=t_2} & \cdots & \left.\frac{\partial \mathbf{y}_R}{\partial \theta_P}\right|_{t=t_2} \end{pmatrix}$$

Yao *et al.*, *Polym. React. Eng.*, 2003

Practical IA

- A parameter is practically identifiable when its **sign** can be determined at the prescribed **confidence level**

Nonlinear Regression $\mathbf{y} = g\left(\mathbf{X};\boldsymbol{\theta}\right) + \boldsymbol{\varepsilon}$

$$\min_{\boldsymbol{\theta}} \left(\mathbf{y} - \mathbf{y}_{obs}\right)^T \left(\mathbf{y} - \mathbf{y}_{obs}\right) = \Phi\left(\hat{\boldsymbol{\theta}}\right)$$

subject to

$$\dot{X}_i = \sum_k \gamma_{ik} \prod_{j=1}^{n+m} X_j^{f_{ijk}}$$

$$\mathbf{y} = g\left(\mathbf{X};\boldsymbol{\theta}\right)$$

$$\boldsymbol{\theta}_l \le \boldsymbol{\theta} \le \boldsymbol{\theta}_u$$

$$g\left(\mathbf{y},\boldsymbol{\theta}\right) \le 0$$

$$h\left(\mathbf{y},\boldsymbol{\theta}\right) = 0$$

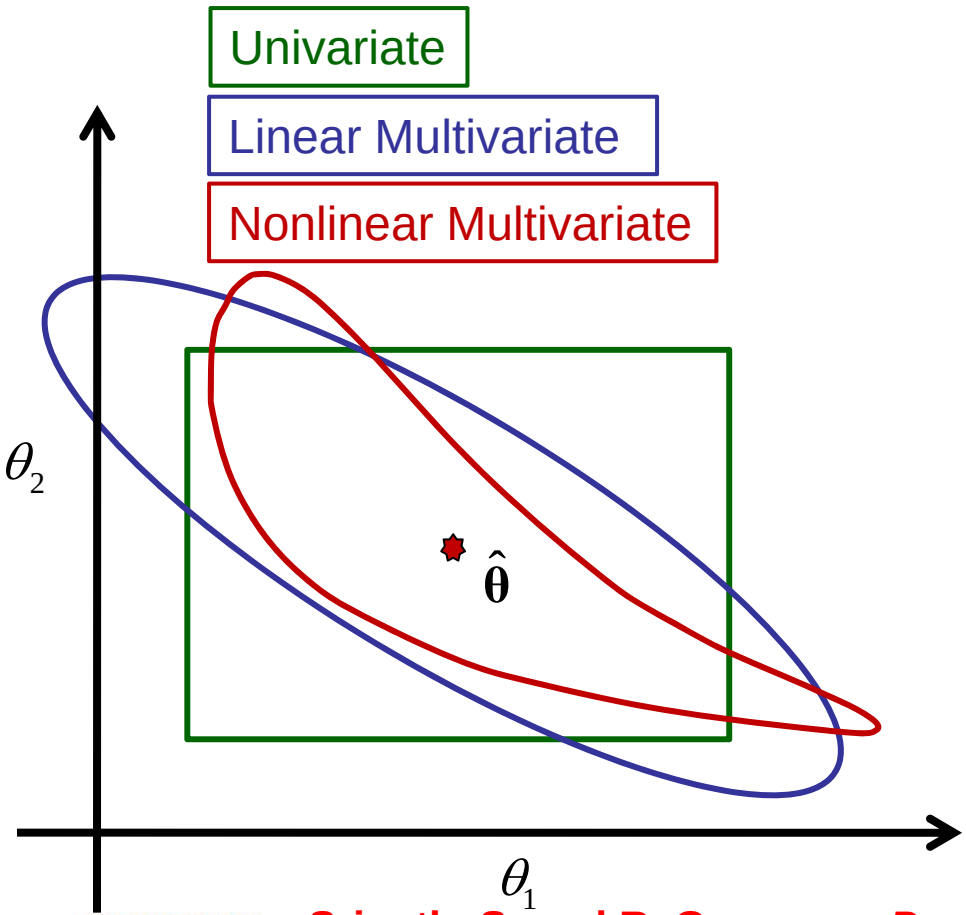
Parameter Confidence Region

If ε is i.i.d $N(0,\sigma^2)$, then

$$\frac{\Phi\left(\boldsymbol{\theta}\right) - \Phi\left(\hat{\boldsymbol{\theta}}\right)}{\Phi\left(\hat{\boldsymbol{\theta}}\right)} \le \frac{p}{n-p} F_{p,n-p}^{\alpha}$$

[Seber and Wild, *Wiley*, 1989]

Practical IA



- Practical identifiability test of a parameter reduces to checking if the **confidence region** crosses $\theta_i = 0$ axis
- Univariate region often under-estimates parameter uncertainty region
- Analysis using linearized model can be misleading

Srinath, S. and R. Gunawan, *Parameter identifiability of power-law biochemical system models. J Biotechnol, 2010*

Practical IA

- Given the parametric uncertainty region, how do we check practical identifiability?

In a linear model or linearized analysis

$$\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\right)^T \mathbf{V}_\theta^{-1}\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\right) \le ps^2 F_{p,n-p}^{\alpha}$$

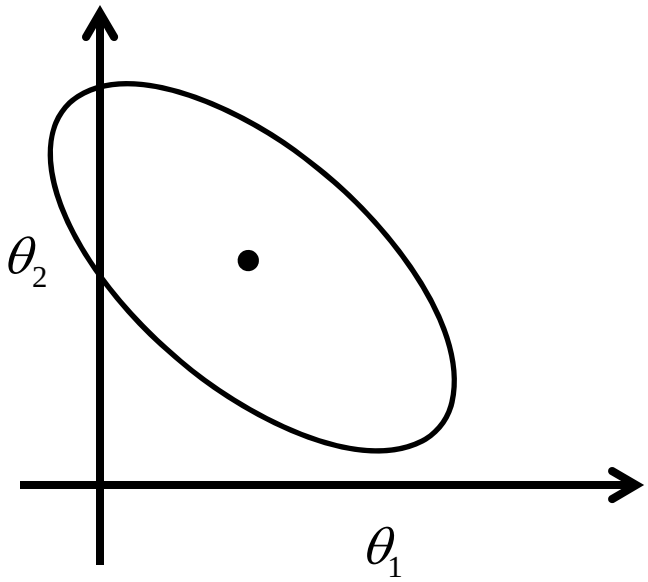
Set $\theta_i = 0$ and recast into feasibility problem of LMI*

$$\begin{bmatrix} -\hat{\theta}_1 \\ \boldsymbol{\theta}_{\setminus 1} - \hat{\boldsymbol{\theta}}_{\setminus 1} \end{bmatrix}^T \begin{bmatrix} \lambda_1 & \boldsymbol{\lambda}_{\setminus 1}^T \\ \boldsymbol{\lambda}_{\setminus 1} & \boldsymbol{\Lambda}_{\setminus 1} \end{bmatrix} \begin{bmatrix} -\hat{\theta}_1 \\ \boldsymbol{\theta}_{\setminus 1} - \hat{\boldsymbol{\theta}}_{\setminus 1} \end{bmatrix} < ps^2 F_{p,n-p}^{\alpha}$$

$$\Rightarrow \Psi + 2\hat{\theta}_1 \boldsymbol{\lambda}_{\setminus 1}^{\mathrm{T}} \boldsymbol{\theta}_{\setminus 1} + 2\hat{\boldsymbol{\theta}}_{\setminus 1} \boldsymbol{\Lambda}_{\setminus 1} \boldsymbol{\theta}_{\setminus 1} - \hat{\boldsymbol{\theta}}_{\setminus 1} \boldsymbol{\Lambda}_{\setminus 1} \boldsymbol{\theta}_{\setminus 1} > 0.$$

$$\Psi = \lambda_1 \hat{\theta}_1^2 + 2\boldsymbol{\lambda}_{\setminus 1}^T \hat{\boldsymbol{\theta}}_{\setminus 1} \hat{\theta}_1 + \hat{\boldsymbol{\theta}}_{\setminus 1}^T \boldsymbol{\Lambda}_{\setminus 1} \hat{\boldsymbol{\theta}}_{\setminus 1} - ps^2 F_{p,n-p}^{\alpha}$$

$$\begin{pmatrix} -\Psi + 2\hat{\theta}_1 \boldsymbol{\lambda}_{\setminus 1}^T \boldsymbol{\theta}_{\setminus 1} + 2\hat{\boldsymbol{\theta}}_{\setminus 1}^T \boldsymbol{\Lambda}_{\setminus 1} \boldsymbol{\theta}_{\setminus 1} & \boldsymbol{\theta}_{\setminus 1}^T \\ \boldsymbol{\theta}_{\setminus 1} & \boldsymbol{\Lambda}_{\setminus 1}^{-1} \end{pmatrix} > 0.$$



*Linear Matrix Inequality

Practical IA

- Given the parametric uncertainty region, how do we check practical identifiability?

Based on nonlinear regression analysis

$$\min_{\boldsymbol{\theta}} \operatorname{sign}(\theta_i)\theta_i$$

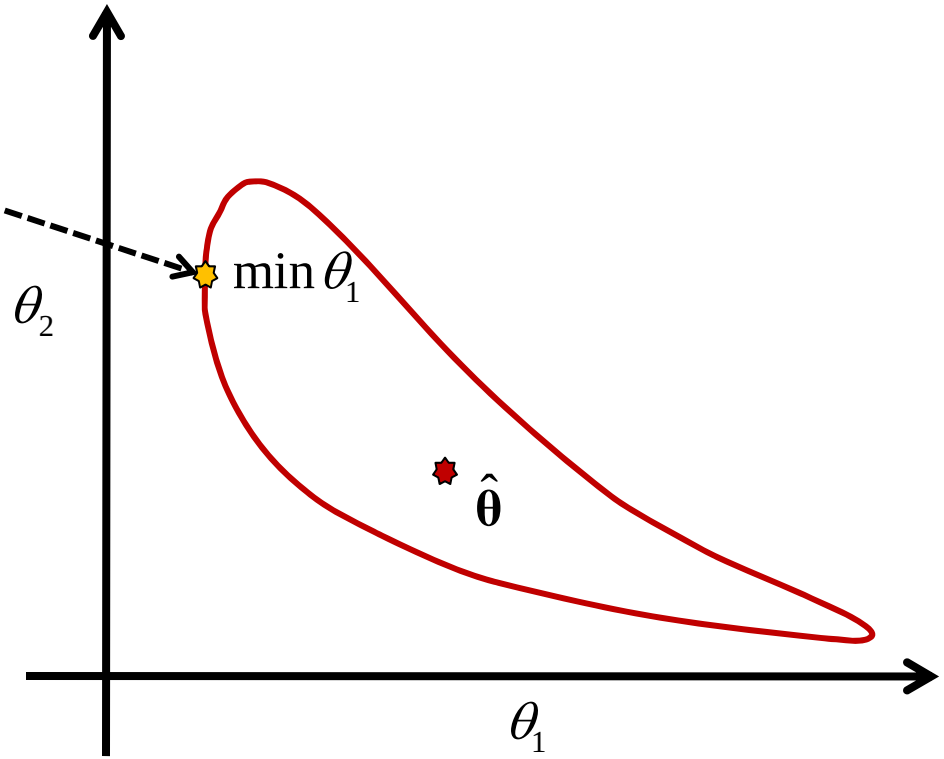
subject to

$$\frac{\Phi(\boldsymbol{\theta}) - \Phi(\hat{\boldsymbol{\theta}})}{\Phi(\hat{\boldsymbol{\theta}})} \le \frac{p}{n-p} F^{\alpha}_{p,n-p}$$

$$\boldsymbol{\theta}_l \le \boldsymbol{\theta} \le \boldsymbol{\theta}_u$$

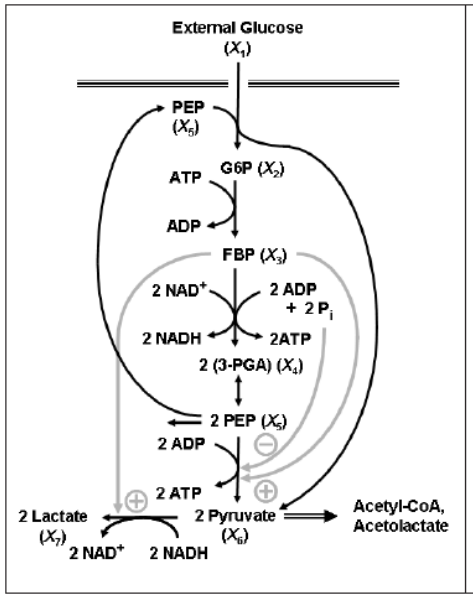
$$g(\mathbf{y},\boldsymbol{\theta}) \le 0$$

$$h(\mathbf{y},\boldsymbol{\theta}) = 0$$



Case Studies

L. lactis model [Voit *et al.*, IET Syst. Biol., 2006]



$$\dot{X}_1 = -\beta_1 X_1^{h_{11}} X_2^{h_{12}} X_5^{h_{25}}$$

$$\dot{X}_2 = \alpha_2 X_1^{h_{11}} X_2^{h_{12}} X_5^{h_{25}} - \beta_2 X_2^{h_{22}} ATP^{h_{2,ATP}}$$

$$\dot{X}_3 = \beta_2 X_2^{h_{22}} ATP^{h_{2,ATP}} - \beta_3 X_3^{h_{33}} P_i^{h_{3,Pi}} NAD^{h_{3,NAD}}$$

$$\dot{X}_4 = 2\beta_3 X_3^{h_{33}} P_i^{h_{3,Pi}} NAD^{h_{3,NAD}} + \alpha_4 X_5^{g_{45}} - \beta_4 X_4^{h_{44}}$$

$$\dot{X}_5 = \beta_4 X_4^{h_{44}} - \alpha_2 X_1^{h_{11}} X_2^{h_{12}} X_5^{h_{25}} - \alpha_4 X_5^{g_{45}} - \beta_{51} X_3^{h_{513}} X_5^{h_{515}} P_i^{h_{51,Pi}} - \beta_{52} X_5^{h_{525}}$$

$$\dot{X}_6 = \alpha_2 X_1^{h_{11}} X_2^{h_{12}} X_5^{h_{25}} + \beta_{51} X_3^{h_{513}} X_5^{h_{515}} P_i^{h_{51,Pi}} - \beta_{61} X_6^{h_{616}} X_3^{h_{613}} NAD^{h_{61,NAD}} - \beta_{62} X_6^{h_{626}}$$

$$\dot{X}_7 = \beta_{61} X_6^{h_{616}} X_3^{h_{613}} NAD^{h_{61,NAD}}$$

States = 6
Parameters = 25

E. coli model [Ko *et al.*, Biochem. Eng. J., 2006]

$$\frac{dX_1}{dt} = \alpha_1 X_1^{g11} X_2^{g12} - \beta_1 X_1^{h11} X_2^{h12}$$

$$\frac{dX_2}{dt} = \alpha_2 - \beta_2 X_1^{h21} X_2^{h22}$$

$$\frac{dX_3}{dt} = \alpha_3 X_1^{g31} X_2^{g32} - \beta_3 X_1^{h31} X_2^{h32} X_3^{h33}$$

$$\frac{dX_4}{dt} = \alpha_4 X_1^{g41} X_2^{g42} - \beta_4 X_1^{h41} X_2^{h42} X_4^{h44}$$

$$\frac{dX_5}{dt} = \alpha_5 X_1^{g51} X_2^{g52} - \beta_5 X_1^{h51} X_2^{h52} X_5^{h55}$$

States = 5
Parameters = 31

Summary of IA results

***L. lactis* Model**					
	Total Parameters	AIP	Practical Identifiability		
			Nonlinear Multivariate	Linear Multivariate	Univariate
Rate constant	9	4	–	0	0
Kinetic order	16	11	–	5	5
***E. coli* Model**					
Rate constant	10	5	5	2	5
Kinetic order	21	10	1	5	6

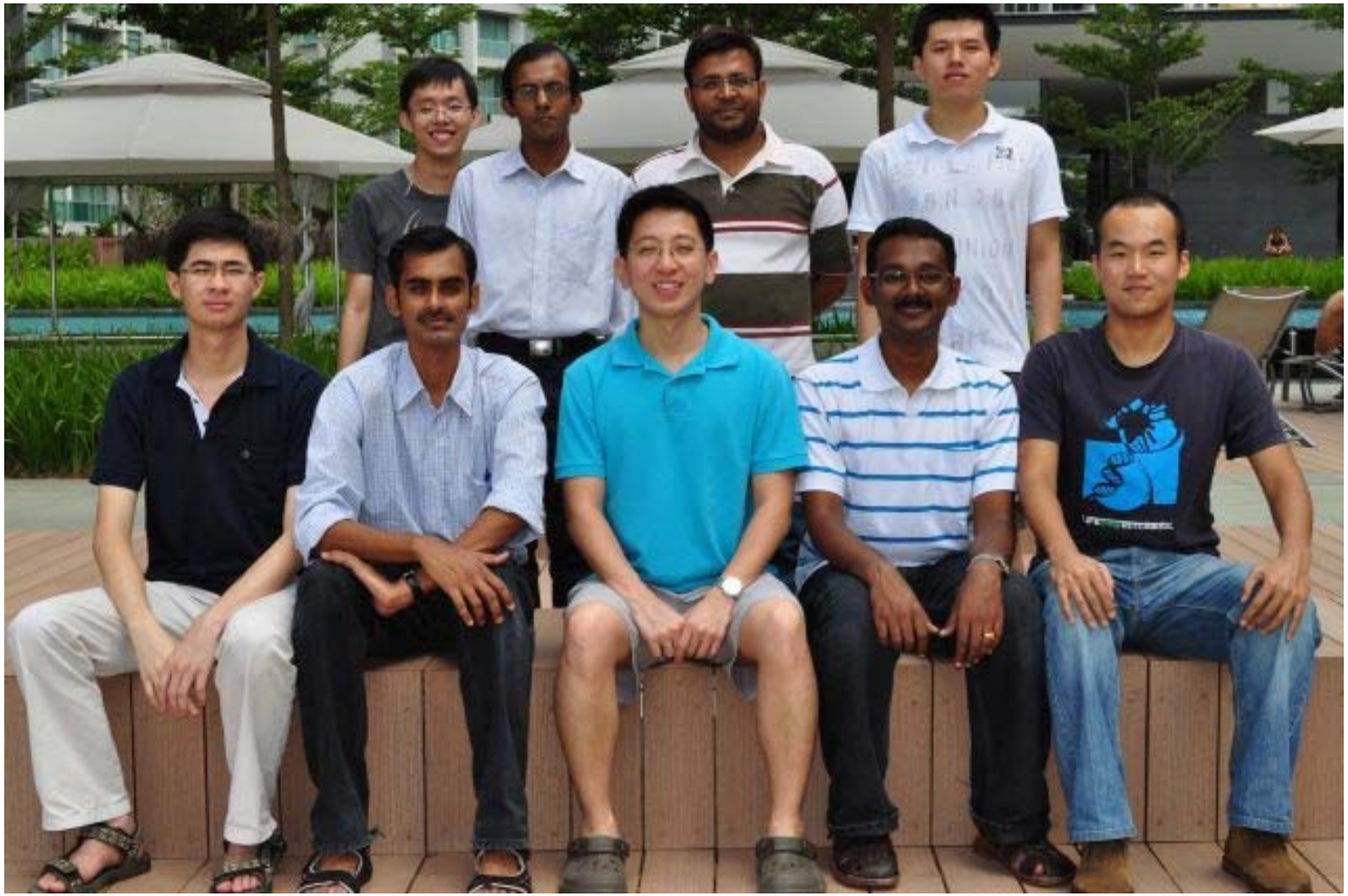
Conclusions

- The inverse modeling problem in biochemical systems is challenging even when using state-of-the-art experimental data.

- The results also suggest that most parameters cannot be uniquely estimated regardless of the estimation algorithms used.

- Addressing these identifiability issues upfront can improve the subsequent reverse engineering of metabolic networks.

Acknowledgements



THANK YOU

(srinath@nus.edu.sg)

Srinath, S. and R. Gunawan, *Parameter identifiability of power-law biochemical system models. J Biotechnol, 2010*