Model discrimination in Systems Biology

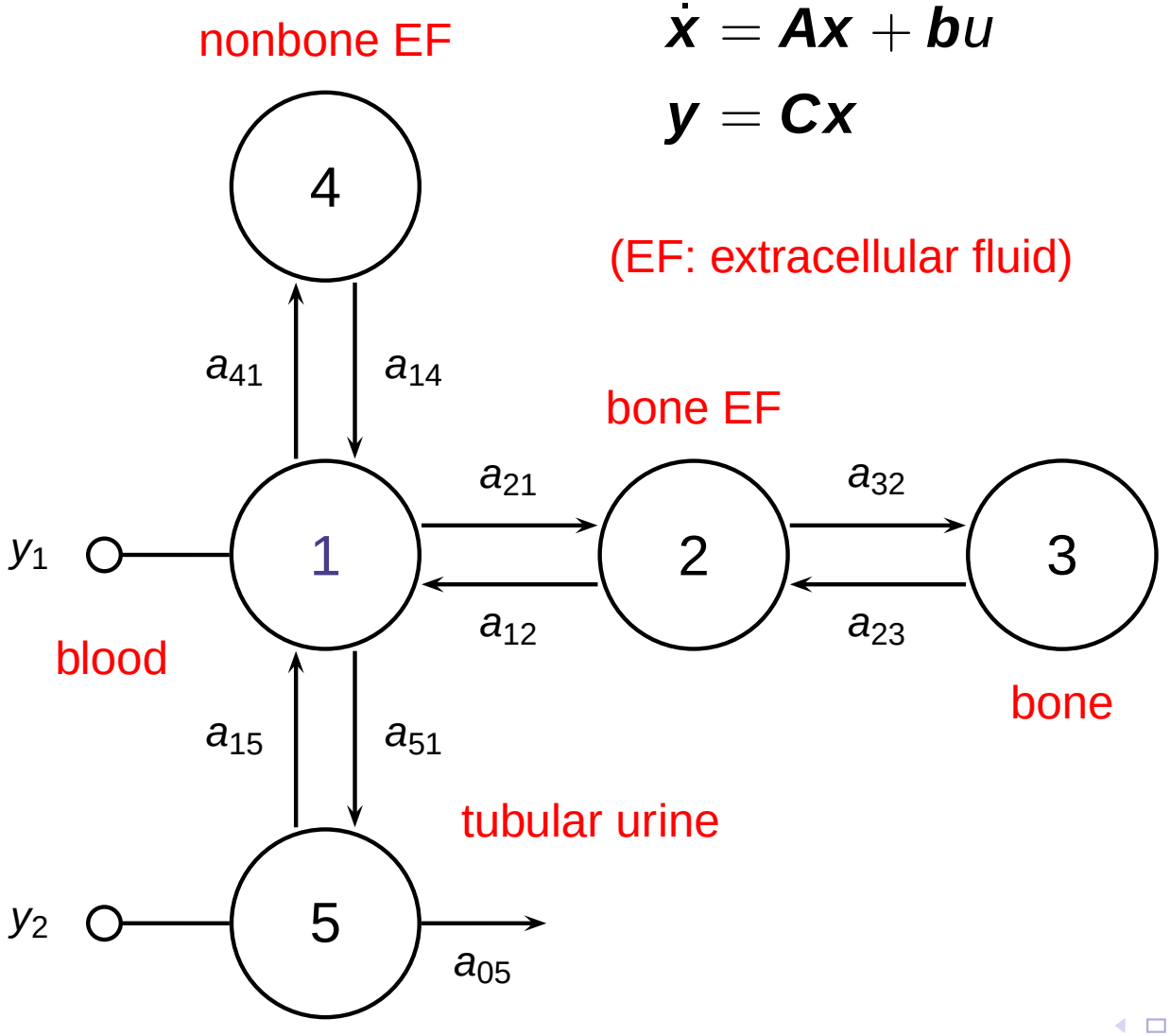
Mike Chappell & Neil Evans
m.j.chappell@warwick.ac.uk

Systems, Measurement and Modelling Research Group
School of Engineering, University of Warwick UK

EPSRC-funded Workshop on Indistinguishability and Model Discrimination in Systems Biology, 29 September 2009, University of Warwick

Outline

Skeletal tracer kinetics model



$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}u$$

$$\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x}$$

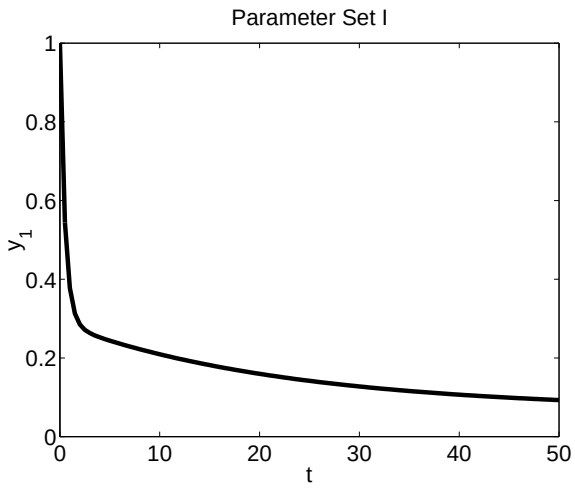
(EF: extracellular fluid)

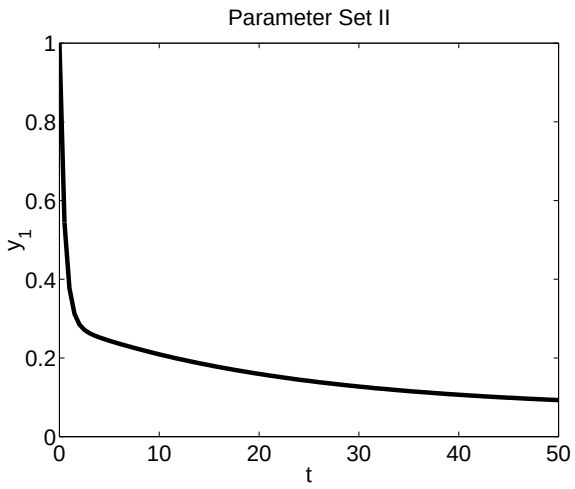
$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & 0 & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & 0 & 0 \\ 0 & a_{32} & -a_{23} & 0 & 0 \\ a_{41} & 0 & 0 & -a_{14} & 0 \\ a_{51} & 0 & 0 & 0 & a_{55} \end{pmatrix}$$

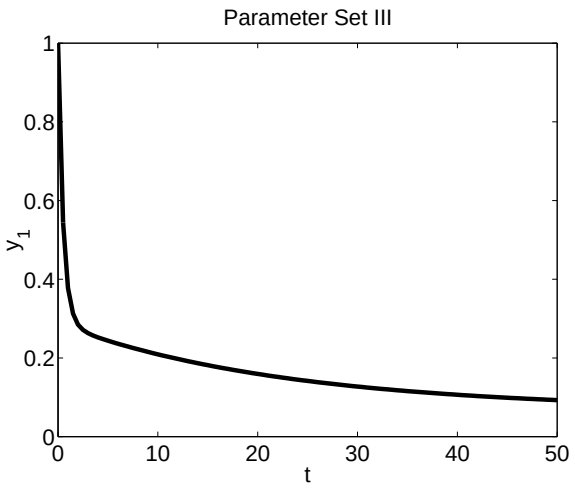
$$a_{ii} = -\sum_{j=0, i\neq j}^{5} a_{ji}$$

	I	II	III
a_{05}	0.612	0.612	0.612
a_{12}	0.908	0.524	0.671
a_{14}	0.567	1.518	0.012
a_{15}	0.388	0.388	0.388
a_{21}	0.246	1.291	1.337
a_{23}	0.020	0.013	1.283
a_{32}	0.602	0.042	0.131
a_{41}	1.191	0.146	0.100
a_{51}	0.024	0.024	0.024

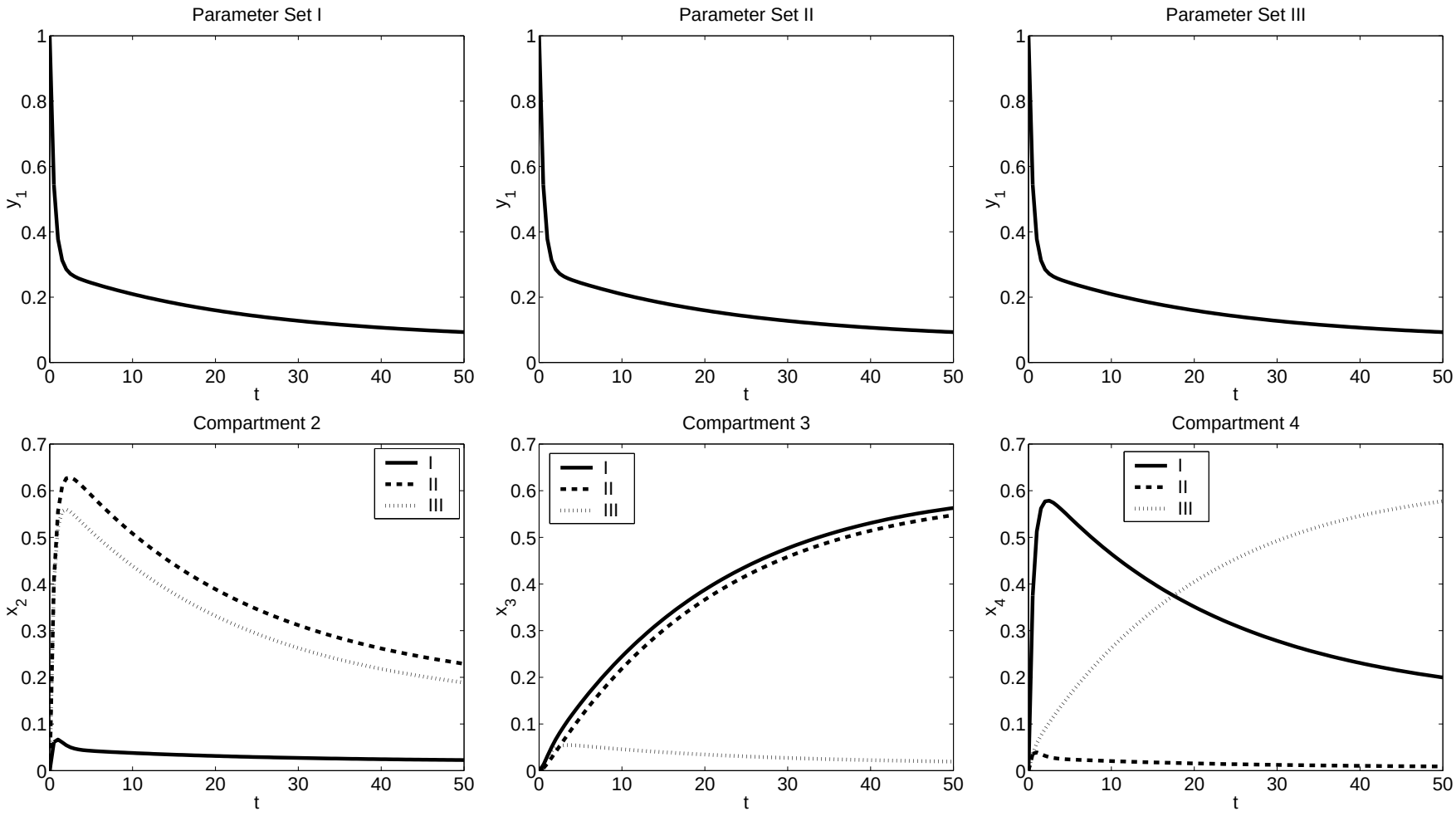
Model simulations





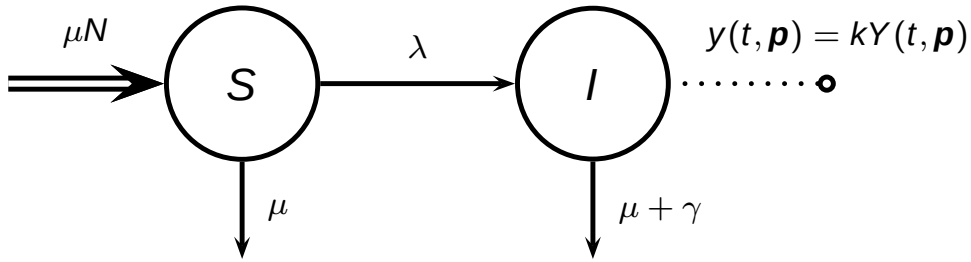


Model simulations



SIR Model

SIR infectious disease model:



Proportion of prevalence measured: $y(t, \boldsymbol{p}) = kY(t, \boldsymbol{p})$

Model equations:

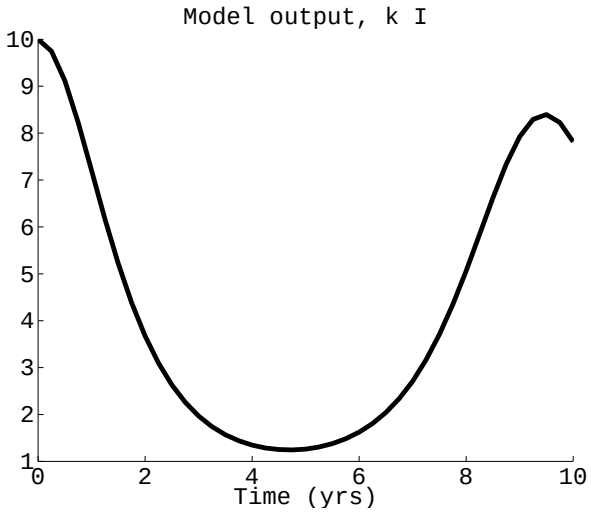
$$\dot{X} = \mu N - \mu X - \frac{\beta}{N} X Y$$

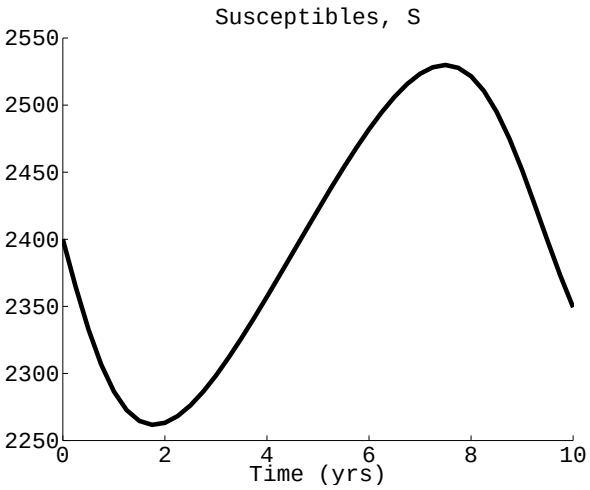
$$\dot{Y} = \frac{\beta}{N} X Y - (\mu + \gamma) Y$$

$$y = kY$$

SIR model

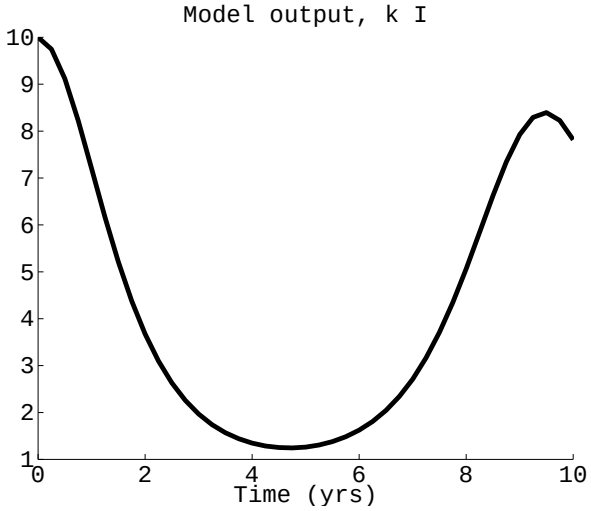
$\mu = 0.0125, \gamma = 12$
$N = 10000$
$\beta = 50, k = 0.5$
$X(0) = 2400$
$Y(0) = 20$

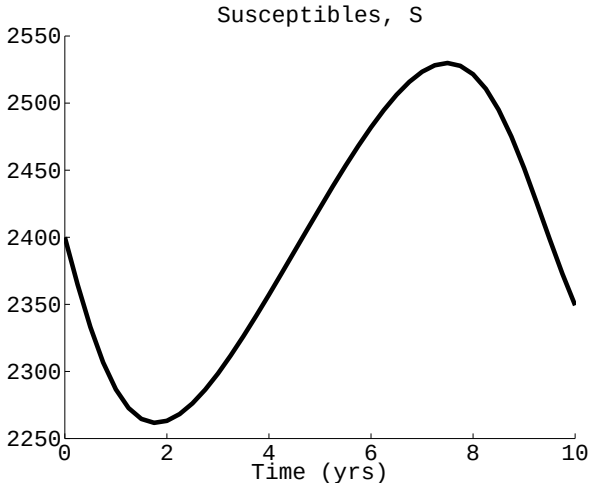




SIR model

$\mu = 0.0125, \gamma = 12$
$N = 10000$
$\beta = 50, k = 0.5$
$X(0) = 2400$
$Y(0) = 20$

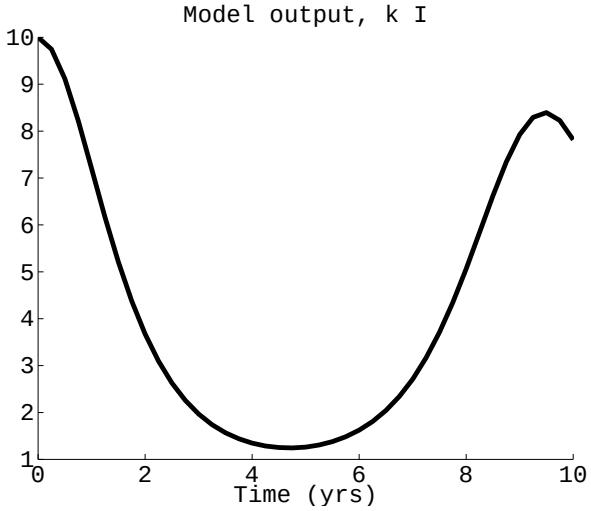


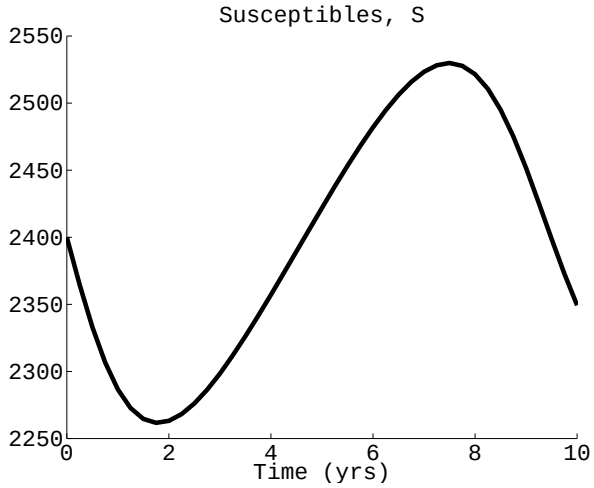


$\mu = 0.0125, \gamma = 12$
$N = 20000$
$\beta = 50, k = 0.25$
$X(0) = 4800$
$Y(0) = 40$

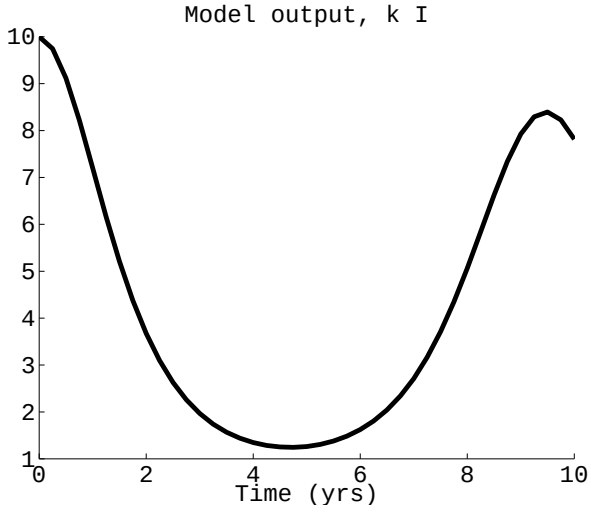
SIR model

$\mu = 0.0125, \gamma = 12$
$N = 10000$
$\beta = 50, k = 0.5$
$X(0) = 2400$
$Y(0) = 20$



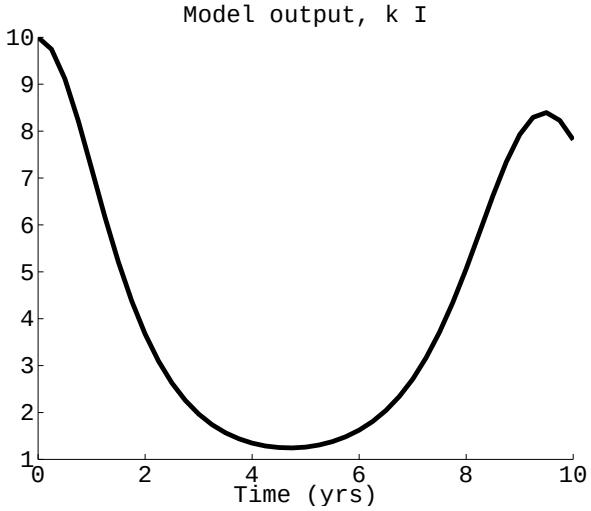


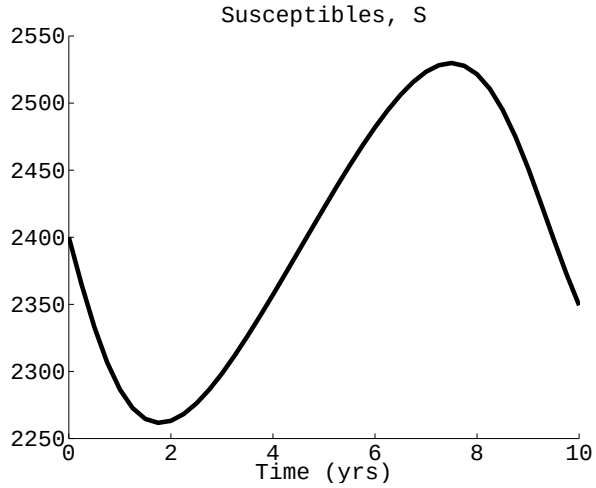
$\mu = 0.0125, \gamma = 12$
$N = 20000$
$\beta = 50, k = 0.25$
$X(0) = 4800$
$Y(0) = 40$



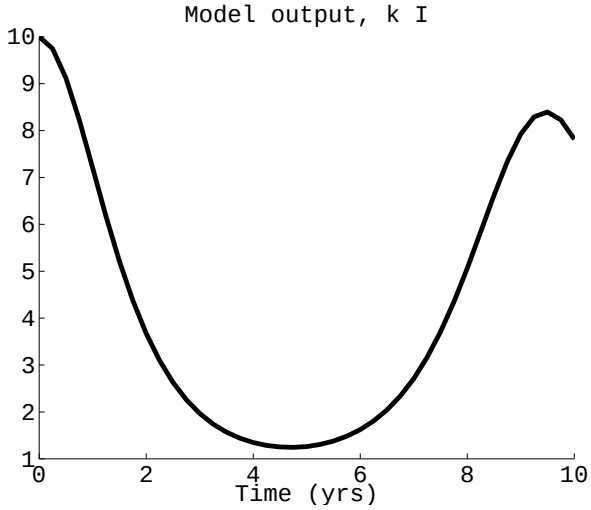
SIR model

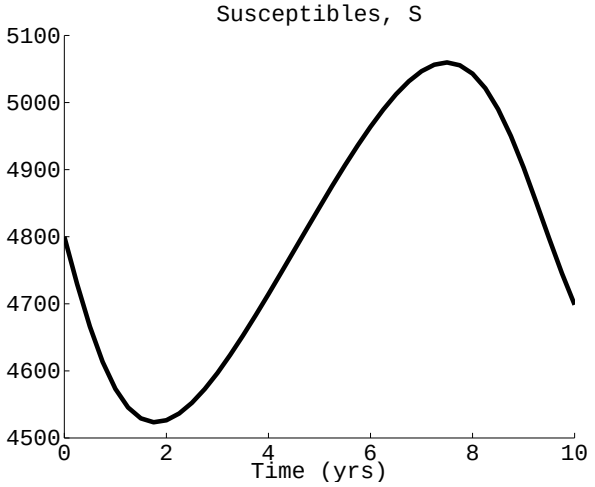
$\mu = 0.0125, \gamma = 12$
$N = 10000$
$\beta = 50, k = 0.5$
$X(0) = 2400$
$Y(0) = 20$



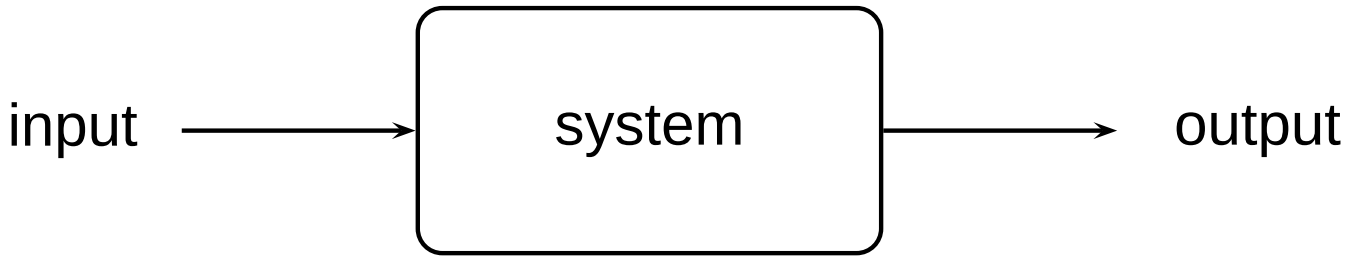


$\mu = 0.0125, \gamma = 12$
$N = 20000$
$\beta = 50, k = 0.25$
$X(0) = 4800$
$Y(0) = 40$





Structural identifiability



Given postulated state-space model, are the unknown parameters uniquely determined by the output (ie, perfect, continuous, noise-free data)?

Necessary theoretical prerequisite to:

- experiment design
- system identification
- parameter estimation

Formal definition

Consider following general parameterised state-space model:

$$\dot{\boldsymbol{x}}(t,\boldsymbol{p}) = \boldsymbol{f}(\boldsymbol{x}(t,\boldsymbol{p}),\boldsymbol{u}(t),\boldsymbol{p}), \quad \boldsymbol{x}(0,\boldsymbol{p}) = \boldsymbol{x}_0(\boldsymbol{p}),$$
$$\boldsymbol{y}(t,\boldsymbol{p}) = \boldsymbol{h}(\boldsymbol{x}(t,\boldsymbol{p}),\boldsymbol{p}),$$

where $\boldsymbol{p}$ is the r-dimensional **vector of unknown parameters**, and is assumed to lie in a set of feasible vectors: $\boldsymbol{p} \in \Omega$.

n dimensional vector $\boldsymbol{x}(t,\boldsymbol{p})$ is state vector, such that $\boldsymbol{q}_0(\boldsymbol{p})$ is the initial state (may depend on the unknown parameters)

m dimensional vector $\boldsymbol{u}(t)$ is input/control vector (our influence on system); what inputs are available depends on experiment to be performed, so $\boldsymbol{u}(\cdot) \in \mathcal{U}$, a **set of admissible inputs** (might be empty).

$\boldsymbol{y}(t,\boldsymbol{p})$ is the l-dimensional output/observation vector (what we can measure in the system). In the following we make explicit that output $\boldsymbol{y}$ depends on $\boldsymbol{p} \in \Omega$ and $\boldsymbol{u} \in \mathcal{U}$ by writing $\boldsymbol{y}(t,\boldsymbol{p};\boldsymbol{u})$.

Parameter identifiability

For generic $\boldsymbol{p} \in \Omega$, the parameter p_i is said to be locally identifiable if there exists a neighbourhood of vectors around $\boldsymbol{p}$, $\mathcal{N}(\boldsymbol{p})$, such that if $\overline{\boldsymbol{p}} \in \mathcal{N}(\boldsymbol{p}) \subseteq \Omega$ and:

$$\text{for every input } \boldsymbol{u} \in \mathcal{U} \text{ and } t \geq 0, \quad \boldsymbol{y}(t, \boldsymbol{p}; \boldsymbol{u}) = \boldsymbol{y}(t, \overline{\boldsymbol{p}}; \boldsymbol{u})$$

then $\overline{p}_i = p_i$.

In particular, if the neighbourhood $\mathcal{N}(\boldsymbol{p}) = \Omega$ can be used in the previous definition, then the parameter p_i is globally/uniquely identifiable.

If the parameter p_i is **not locally identifiable**, i.e., there is no suitable neighbourhood $\mathcal{N}(\boldsymbol{p})$, then it is said to be unidentifiable.

Structural identifiability

Structurally globally/uniquely identifiable

A compartmental model is structurally globally/uniquely identifiable (**SGI**) if all of the unknown parameters p_i are globally/uniquely identifiable.

Structurally locally identifiable

A compartmental model is structurally locally identifiable (**SLI**) if all of the unknown parameters p_i are locally identifiable and at least one of these parameters is **not** globally identifiable.

Unidentifiable

A compartmental model is unidentifiable if at least one of the unknown parameters p_i is unidentifiable.

Remarks

- Necessary condition for parameter estimation
  - Essential for parameters with practical significance
  - Prerequisite to experiment design
- Identifiability does not guarantee
  - Good fit to experimental data
  - Good fit only with unique vector of parameters
- Unidentifiable implies infinite number of parameter vectors will give same fit (even for perfect data)
- Many techniques for linear systems
  - Laplace transform or transfer function
  - Taylor series of output
  - Similarity transformation (exhaustive modelling)
- Taylor series and similarity transformation approaches are applicable for nonlinear systems
- Differential algebra
  - Polynomial systems with differentiable inputs/outputs
  - Heavily dependent on symbolic computation

General linear system

$$\dot{\boldsymbol{x}}(t,\boldsymbol{p}) = \boldsymbol{A}(\boldsymbol{p})\boldsymbol{x}(t,\boldsymbol{p}) + \boldsymbol{B}(\boldsymbol{p})\boldsymbol{u}(t), \quad \boldsymbol{x}(0,\boldsymbol{p}) = \boldsymbol{x}_0(\boldsymbol{p}),$$
$$\boldsymbol{y}(t,\boldsymbol{p}) = \boldsymbol{C}(\boldsymbol{p})\boldsymbol{x}(t,\boldsymbol{p}),$$

where

$\boldsymbol{A}(\boldsymbol{p})$ is an $n \times n$ matrix of rate constants

$\boldsymbol{B}(\boldsymbol{p})$ is an $n \times m$ input matrix

$\boldsymbol{C}(\boldsymbol{p})$ is an $l \times n$ output matrix

Assume that $\boldsymbol{x}_0 = 0$ (not essential) & take Laplace transforms:

$$s\boldsymbol{Q}(s) = \boldsymbol{A}(\boldsymbol{p})\boldsymbol{Q}(s) + \boldsymbol{B}(\boldsymbol{p})\boldsymbol{U}(s)$$
$$\boldsymbol{Y}(s) = \boldsymbol{C}(\boldsymbol{p})\boldsymbol{Q}(s)$$
$$= \boldsymbol{C}(\boldsymbol{p})\left(s\boldsymbol{I}_n - \boldsymbol{A}(\boldsymbol{p})\right)^{-1}\boldsymbol{B}(\boldsymbol{p})\boldsymbol{U}(s)$$

Laplace Transform Approach

This gives relationship between LTs of input & output:

$$\boldsymbol{Y}(s) = \boldsymbol{G}(s)\boldsymbol{U}(s),$$

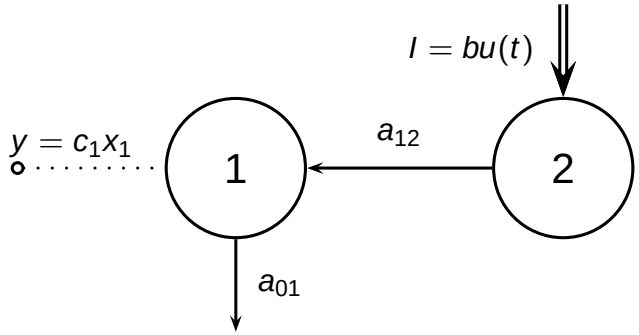
where the matrix

$$\boldsymbol{G}(s) = \boldsymbol{C}(\boldsymbol{p})\left(s\boldsymbol{I}_n - \boldsymbol{A}(\boldsymbol{p})\right)^{-1}\boldsymbol{B}(\boldsymbol{p})$$

is the transfer (function) matrix

- Measurements for $\boldsymbol{G}(s)$ assumed known
- Coefficients of powers of s in numerators & denominators uniquely determined by input-output relationship

Example: 2 Compartments



Model is:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -a_{01} & a_{12} \\ 0 & -a_{12} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ b \end{bmatrix} u(t)$$

$$y = \begin{bmatrix} c & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Transfer function:

$$G(s) = \begin{bmatrix} c & 0 \end{bmatrix} \begin{bmatrix} s + a_{01} & -a_{12} \\ 0 & s + a_{12} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ b \end{bmatrix} = \frac{bca_{12}}{(s + a_{01})(s + a_{12})}$$

A cautionary tale

Transfer function:

$$G(s) = \frac{bca_{12}}{(s + a_{01})(s + a_{12})}$$

and so the following are unique:

$$bca_{12}, \qquad a_{01} + a_{12} \qquad \text{and} \qquad a_{01}a_{12}$$

- Yields two possible solutions for a_{01} and a_{21}
- If b (or c) known then two possible solutions for c (or b) hence locally identifiable
- If neither b nor c known then unidentifiable
- If both b and c known then globally identifiable

Techniques for nonlinear models

Generally more difficult to apply, can be less systematic and do not always yield full information concerning identifiability of given system.

One must also be careful about what inputs there are to the system.

Dealing with models of the form:

$$\dot{\boldsymbol{x}}(t,\boldsymbol{p}) = \boldsymbol{f}(\boldsymbol{x}(t,\boldsymbol{p}),\boldsymbol{p}), \quad \boldsymbol{x}(0,\boldsymbol{p}) = \boldsymbol{x}_0(\boldsymbol{p}),$$
$$\boldsymbol{y}(t,\boldsymbol{p}) = \boldsymbol{h}(\boldsymbol{x}(t,\boldsymbol{p}),\boldsymbol{p}),$$

where

- $\boldsymbol{p} \in \Omega$ is an r dimensional vector
- $\boldsymbol{x}(t,\boldsymbol{p})$ is an n dimensional vector
- $\boldsymbol{y}(t,\boldsymbol{p})$ is an l dimensional vector

Taylor series approach

This approach for linear models also works for nonlinear ones:

$$y_i(t,\boldsymbol{p}) = y_i(0,\boldsymbol{p}) + \dot{y}_i(0,\boldsymbol{p})t + \ddot{y}_i(0,\boldsymbol{p})\frac{t^2}{2!} + \cdots + y_i^{(k)}(0,\boldsymbol{p})\frac{t^k}{k!} + \ldots,$$

where $y_i^{(k)}(0,\boldsymbol{p}) = \left.\dfrac{\mathrm{d}^k y_i}{\mathrm{d}t^k}\right|_{t=0} \qquad (k = 1, 2, \ldots).$

Taylor series coefficients $y_i^{(k)}(0,\boldsymbol{p})$ unique for a particular output

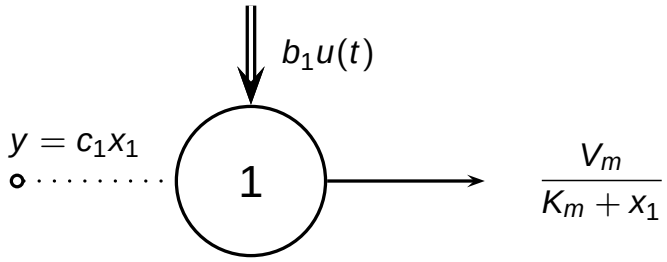
Notice that we have a possibly infinite list of coefficients:

$$y_1(0,\boldsymbol{p}), \ldots, y_l(0,\boldsymbol{p}), \dot{y}_1(0,\boldsymbol{p}), \ldots, \dot{y}_l(0,\boldsymbol{p}), \ddot{y}_1(0,\boldsymbol{p}), \ldots, \ddot{y}_l(0,\boldsymbol{p}), \ldots$$

and upper bound on number of coefficients needed more difficult

It is quite difficult to use the Taylor series approach to prove that a model is unidentifiable.

Example: 1 compartment



Model equations:

$$\dot{x}_1 = -\frac{V_m x_1}{K_m + x_1}, \qquad x_1(0) = b_1$$

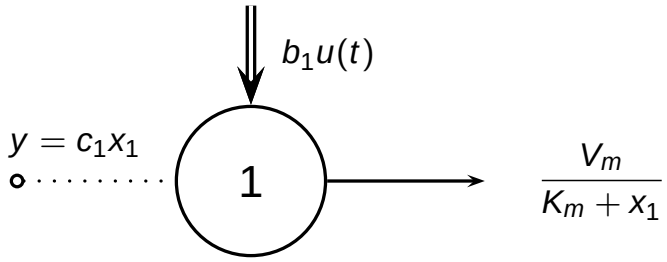
$$y = c_1 x_1$$

First coefficient: $y(0, \boldsymbol{p}) = b_1 c_1$

Second coefficient: $\dot{y}(0, \boldsymbol{p}) = -\dfrac{c_1 V_m b_1}{K_m + b_1}$

Third coefficient: $y^{(2)}(t, \boldsymbol{p}) = \dfrac{\mathrm{d}}{\mathrm{d}t}\left(-\dfrac{c_1 V_m x_1}{K_m + x_1}\right)$

Example: 1 compartment



Model equations:

$$\dot{x}_1 = -\frac{V_m x_1}{K_m + x_1}, \qquad x_1(0) = b_1$$

$$y = c_1 x_1$$

First coefficient: $y(0, \boldsymbol{p}) = b_1 c_1$

Second coefficient: $\dot{y}(0, \boldsymbol{p}) = -\dfrac{c_1 V_m b_1}{K_m + b_1}$

Third coefficient: $y^{(2)}(t, \boldsymbol{p}) = \dfrac{\mathrm{d}}{\mathrm{d}t}\left(-\dfrac{c_1 V_m x_1}{K_m + x_1}\right)$

Horrible! Use MATHEMATICA!

More general problem

Suppose that we wish to model a single substrate, single enzyme reaction—which scheme is appropriate?

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} C \xrightarrow{k_2} P + E \qquad \text{(MM)}$$

$$C \underset{\tilde{k}_{-1}}{\overset{\tilde{k}_1}{\rightleftarrows}} S + E \xrightarrow{\tilde{k}_2} P + E \qquad \text{(H)}$$

- What do we need to measure to find out?

Measuring product only

System equations:

$$\dot{s} = -k_1(e_0 - c)s + k_{-1}c \qquad \dot{\tilde{s}} = -(\tilde{k}_1 + \tilde{k}_2)(\tilde{e}_0 - \tilde{c})\tilde{s} + \tilde{k}_{-1}\tilde{c}$$

$$\dot{c} = k_1(e_0 - c)s - (k_{-1} + k_2)c \qquad \dot{\tilde{c}} = \tilde{k}_1(\tilde{e}_0 - \tilde{c})\tilde{s} - \tilde{k}_{-1}\tilde{c}$$

with $s(0) = s_0$, $\tilde{s}(0) = \tilde{s}_0$ and $c(0) = \tilde{c}(0) = 0$.

Measuring product: $y(t) = \epsilon p(t)$, $\tilde{y}(t) = \tilde{\epsilon}\tilde{p}(t)$.

Comparing terms of Taylor series:

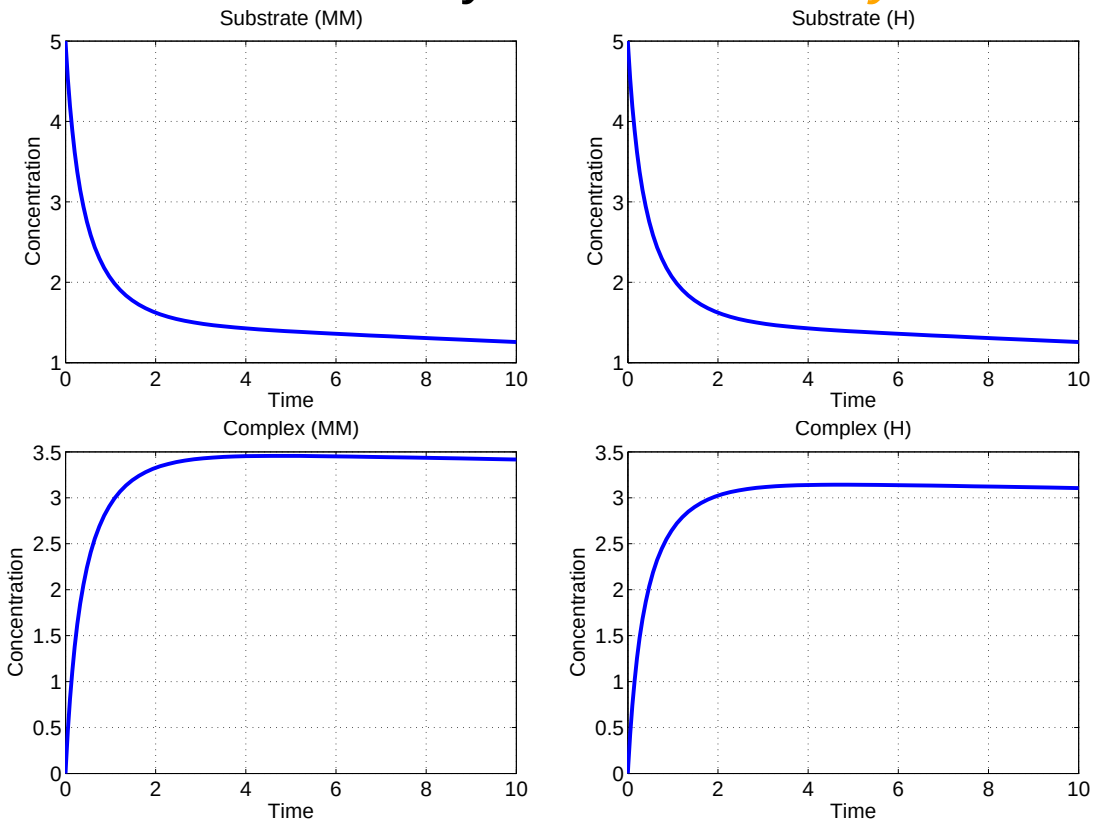
- $y(0) = 0 = \tilde{y}(0)$
- $\dot{y}(0) = \epsilon k_2 c(0) = 0$
- *BUT* $\dot{\tilde{y}}(0) = \tilde{\epsilon}\tilde{k}_2\tilde{s}_0\tilde{e}_0$

Hence we can *distinguish* between the two schemes when measuring product—the outputs are not identical

Measuring substrate only

Outputs are of the form: $y(t) = \epsilon s(t), \tilde{y}(t) = \tilde{\epsilon}\tilde{s}(t)$

It can be shown that the outputs of the two reaction schemes are identical—they are *structurally indistinguishable*



Structural indistinguishability

Consider following pair of systems:

$$\Sigma(\boldsymbol{p}) \begin{cases} \dot{\boldsymbol{x}}(t,\boldsymbol{p}) = \boldsymbol{f}(\boldsymbol{x}(t,\boldsymbol{p}),\boldsymbol{p}), & \boldsymbol{x}(0,\boldsymbol{p}) = \boldsymbol{x}_0(\boldsymbol{p}), \\ \boldsymbol{y}(t,\boldsymbol{p}) = \boldsymbol{h}(\boldsymbol{x}(t,\boldsymbol{p}),\boldsymbol{p}). \end{cases}$$

$$\tilde{\Sigma}(\tilde{\boldsymbol{p}}) \begin{cases} \dot{\tilde{\boldsymbol{x}}}(t,\tilde{\boldsymbol{p}}) = \tilde{\boldsymbol{f}}(\tilde{\boldsymbol{x}}(t,\tilde{\boldsymbol{p}}),\tilde{\boldsymbol{p}}), & \tilde{\boldsymbol{x}}(0,\tilde{\boldsymbol{p}}) = \tilde{\boldsymbol{x}}_0(\tilde{\boldsymbol{p}}), \\ \tilde{\boldsymbol{y}}(t,\tilde{\boldsymbol{p}}) = \tilde{\boldsymbol{h}}(\tilde{\boldsymbol{x}}(t,\tilde{\boldsymbol{p}}),\tilde{\boldsymbol{p}}), \end{cases}$$

- $\Sigma(\boldsymbol{p})$ and $\tilde{\Sigma}(\tilde{\boldsymbol{p}})$ ($\boldsymbol{p} \in \Omega$, $\tilde{\boldsymbol{p}} \in \tilde{\Omega}$) *output indistinguishable*, $\Sigma(\boldsymbol{p}) \sim \tilde{\Sigma}(\tilde{\boldsymbol{p}})$, if $\boldsymbol{y}(t,\boldsymbol{p}) = \tilde{\boldsymbol{y}}(t,\tilde{\boldsymbol{p}})$ for all t.
- Σ and $\tilde{\Sigma}$ *structurally indistinguishable* if
  - for generic $\boldsymbol{p} \in \Omega$ there exists $\tilde{\boldsymbol{p}} \in \tilde{\Omega}$ s.t. $\Sigma(\boldsymbol{p}) \sim \tilde{\Sigma}(\tilde{\boldsymbol{p}})$;
  - for generic $\tilde{\boldsymbol{p}} \in \tilde{\Omega}$ there exists $\boldsymbol{p} \in \Omega$ s.t. $\Sigma(\boldsymbol{p}) \sim \tilde{\Sigma}(\tilde{\boldsymbol{p}})$.

Summary

- Structural identifiability is an important step in modelling process
  - Theoretical prerequisite to experiment design, system identification, and parameter estimation
  - Techniques involve generation, manipulation & solution of nonlinear algebraic equations
  - Need for more tractable techniques for nonlinear systems
- Structural indistinguishability similarly important (more general framework)
- Both are highly relevant to models in Biomedical Systems Modelling/Systems Biology