UNIVERSITÀ
DEGLI STUDI
DI BERGAMO

Dipartimento
di Ingegneria Gestionale,
dell'Informazione e della Produzione

Lesson 12.

Introduction to systems identification

CONTROL AND MODELING OF BIOLOGICAL SYSTEMS

MASTER DEGREE IN MEDICAL ENGINEERING

TEACHER
Antonio Ferramosca

PLACE
University of Bergamo

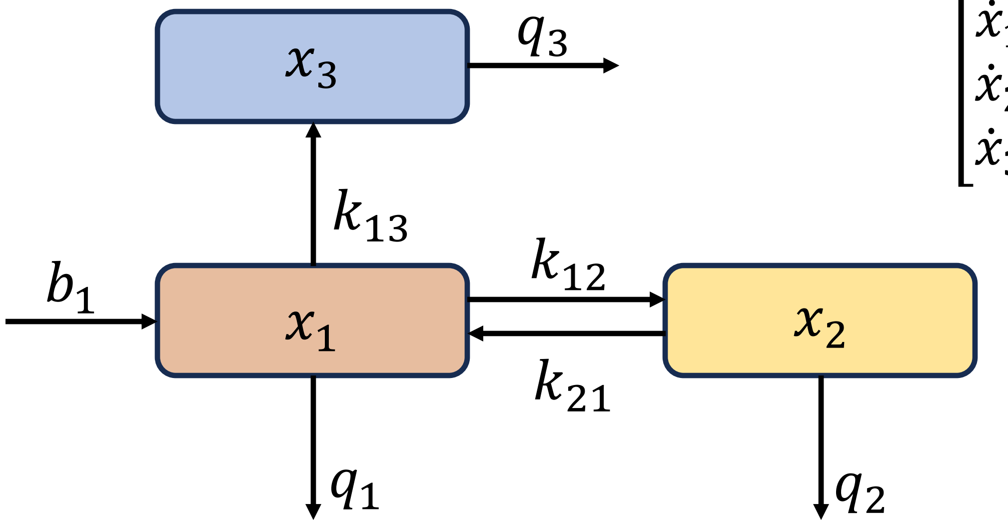
Outline

1. **Introduction to identification**
2. Experiment design and model selection
3. Identification criterion
4. Model validation
5. A priori identifiabilty

What we have seen

So far, we studied:

Modelling: How to model biological systems by means of well-known model structures



$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\\ \dot{x}_3\end{bmatrix} = \begin{bmatrix}-q_1 - k_{12} - k_{13} & k_{21} & 0\\ k_{12} & -q_2 - k_{21} & 0\\ k_{13} & 0 & -q_3\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix} + \begin{bmatrix}b_1\\ 0\\ 0\end{bmatrix}u$$

Given a compartmental structure we know how to obtain the differential equation for such a model

What we have NOT seen

If we do not know the value of the parameters k_{12}, k_{13}, k_{21}, q_1, q_2, q_3, b_1 how can we simulate and test the model?

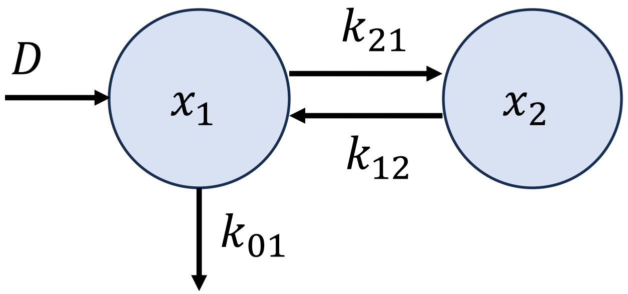
If for instance we have to determine the right posology for a certain drug, how can we make all the simulation tests necessary to test different strategies, if the parameters are unknown?

We need an extra step:

Identification: how do we determine the parameters of a model that describes some phenomenon, based on a certain dataset?

Example

How can we assign values to the parameters of a pharmacokinetic 2-compartments model is of the form:

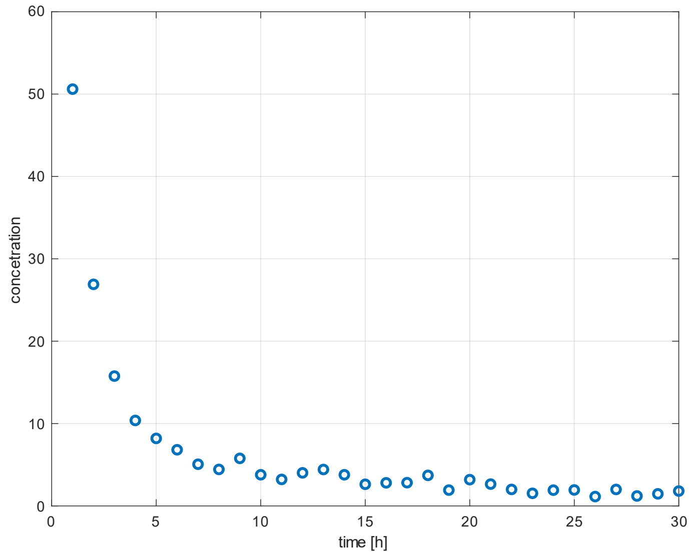


$$\dot{x_1}(t) = -(k_{01} + k_{21})x_1(t) + k_{12}x_2(t)$$

$$\dot{x_2}(t) = k_{21}x_1(t) - k_{12}x_2(t)$$

$$y(t) = \frac{x_1(t)}{V_1}$$

$$D = 500$$



We run an experiment perturbating the system (with a certain input R) and we observe the effect (we measure the output). This measure reflects the values of the parameters of the model

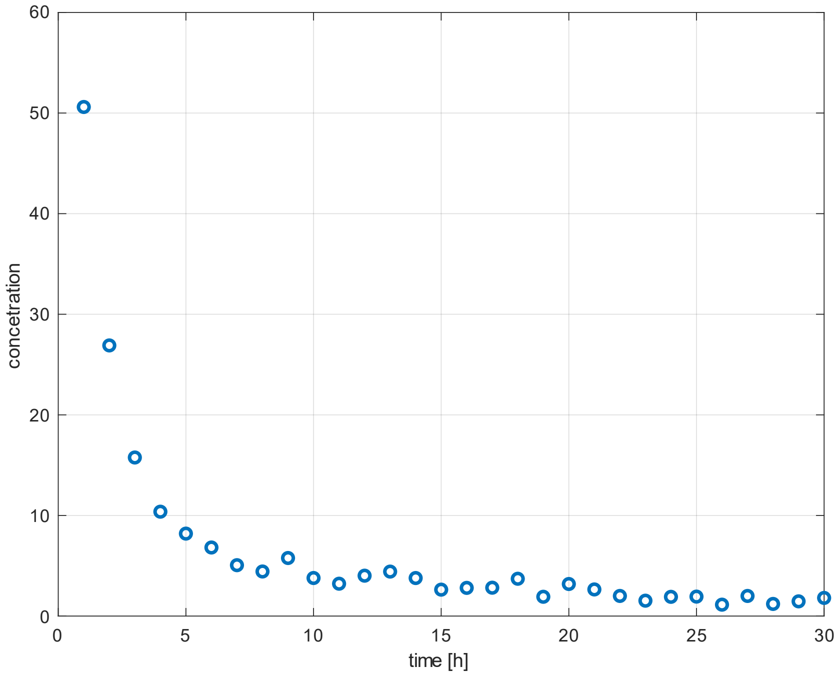
$$y(t) = h(t, k_{01}, k_{12}, k_{21}, V_1)$$

$$= h(t, \boldsymbol{\theta})$$

Example

$$y(t) = h(t, k_{01},\ k_{12},\ k_{21}, V_1)$$
$$= h(t, \boldsymbol{\theta})$$

The value of the output $y(t) = h(t, \boldsymbol{\theta})$ is the prediction of the model (its in continuous time)



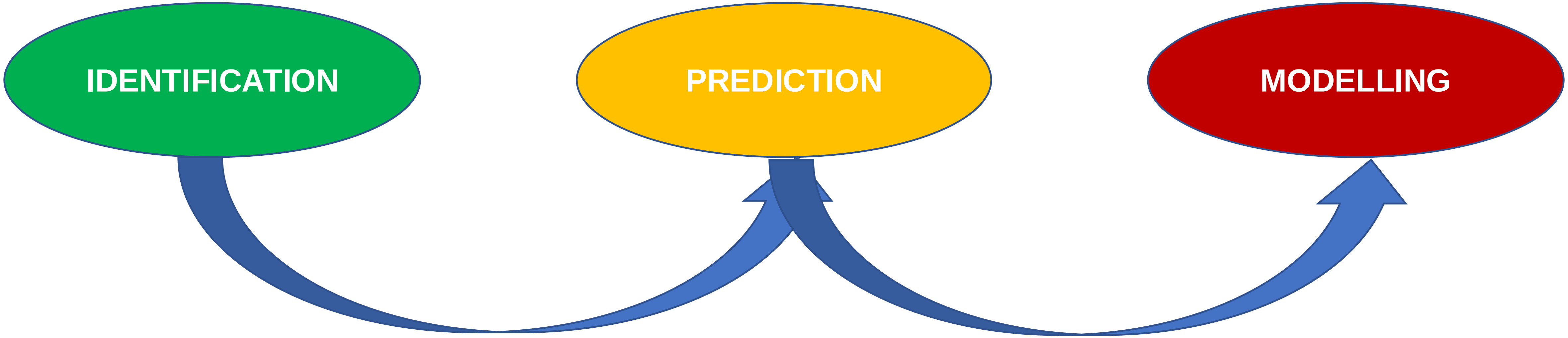
The real output of the experiment, i.e. the measures that we get from the experiment (in discrete time) is affected by some noise of measurement.

What is the goal of identification?

Given a certain dataset

Obtain a good predictor

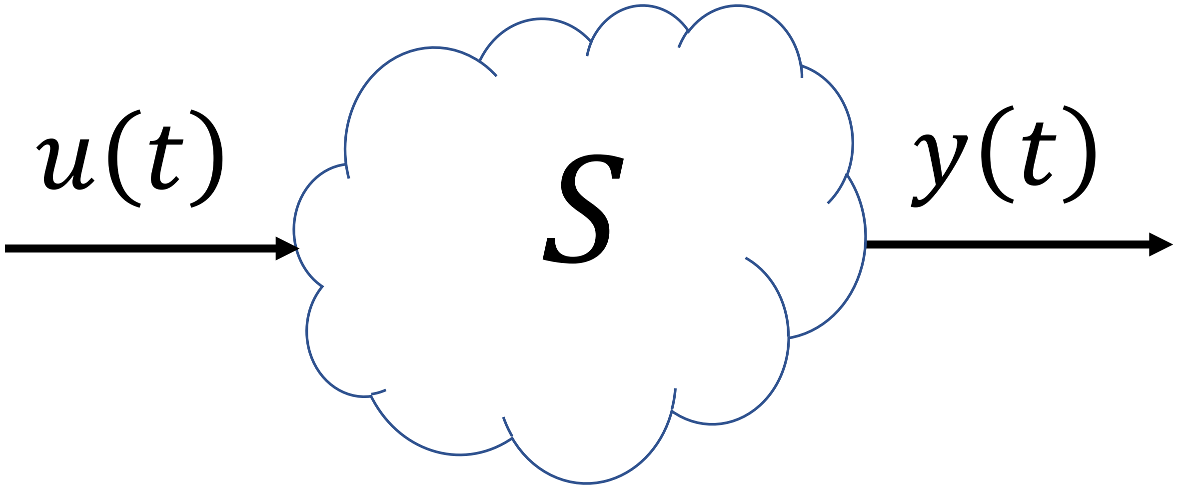
GOAL



Find the optimal parameters

In order to forecast a certain phenomenon

Input/output systems



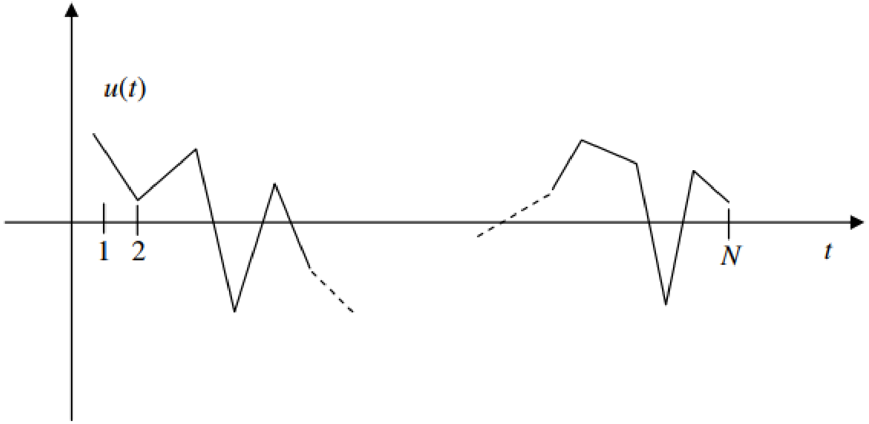
Data generating system

- $y(t)$ is the output variable
- $u(t)$ is the input variable

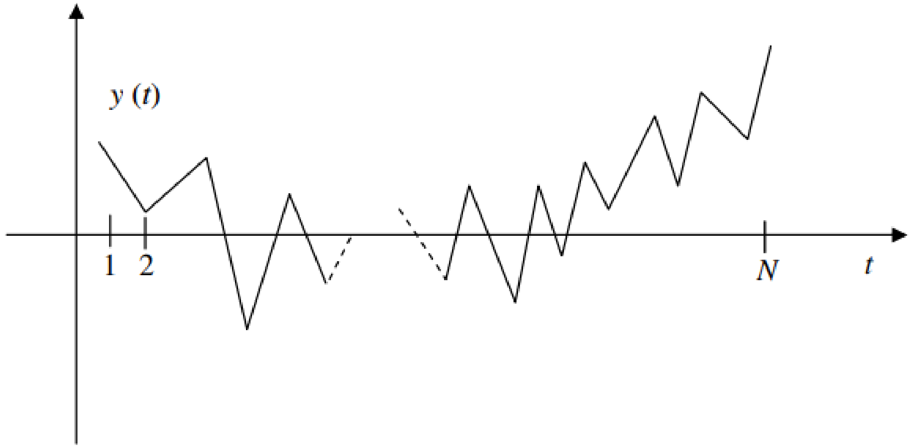
Two sets of N data are collected

$$\{u(1), u(2), \dots, u(N)\}\ \{y(1), y(2), \dots, y(N)\}$$

Input dataset

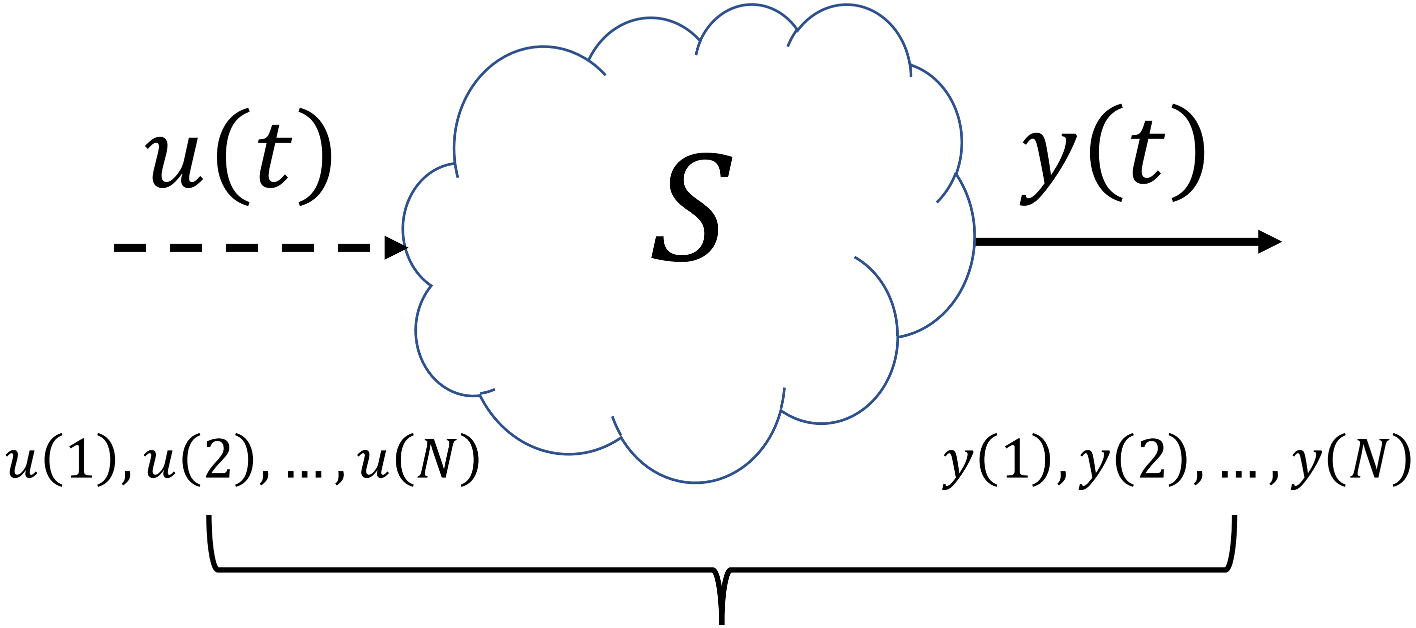


Output dataset



System identification

Retrieve suitable model from **experiments** on the real system

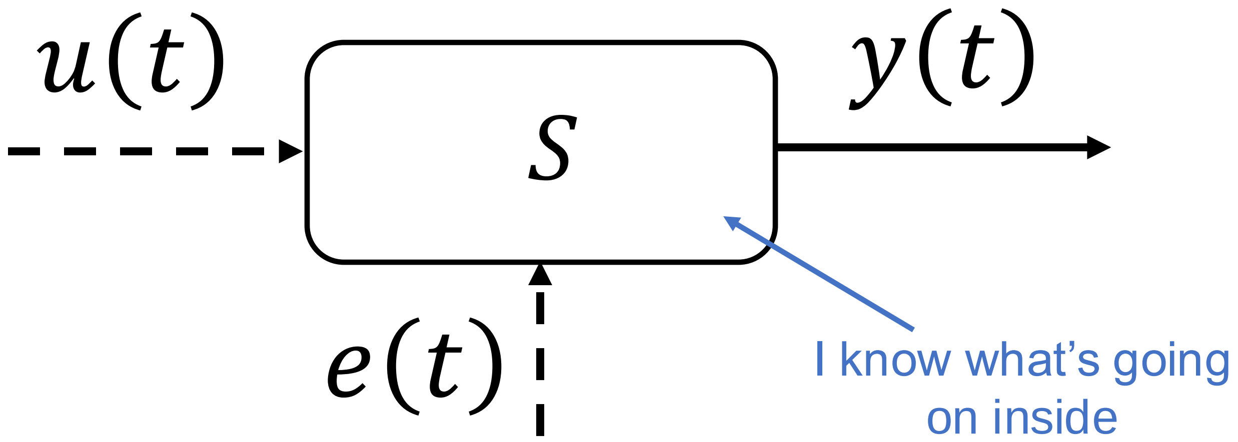


Experiments on the true system

Identification problem: define an **automatic procedure** to find a model for S based on available (input/output or time series) data

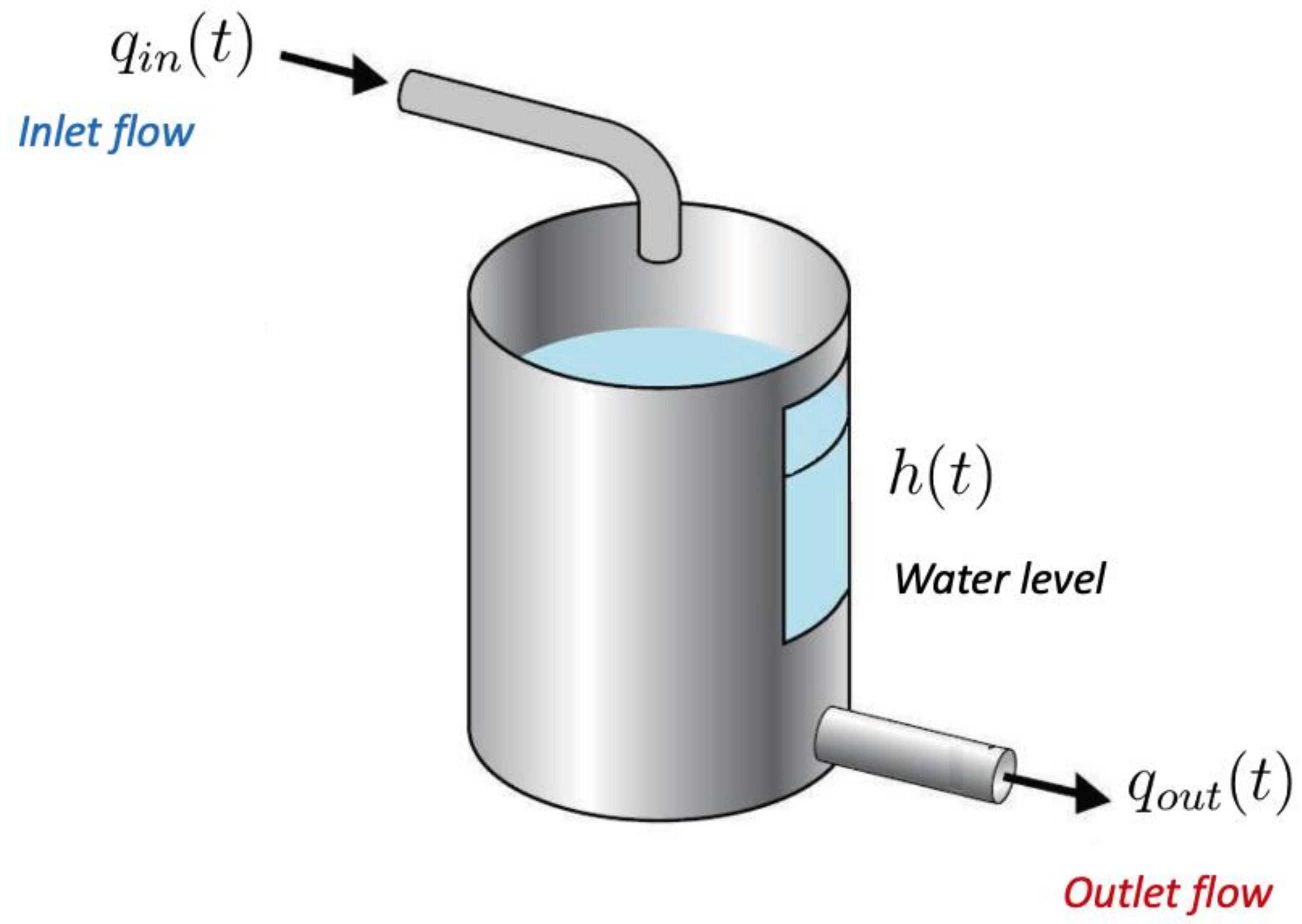
White-box model

White-box model: both the numerical parameters and the functional form of the relation between variables are known



This would be the ideal situation since it implies that we are able to derive an exact mathematical model for the underlying process.

Example: water tank



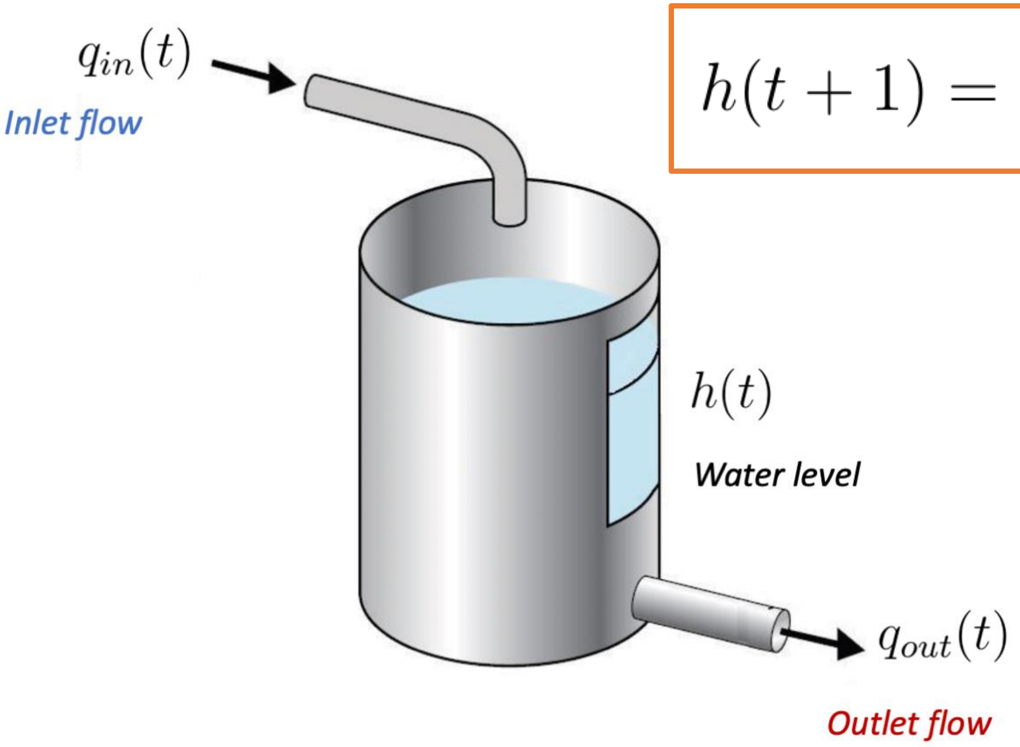
$$q_{in}(t) - q_{out}(t) = A\frac{dh(t)}{dt}$$

$$q_{out}(t) = \kappa\sqrt{h(t)}$$

$$\frac{dh(t)}{dt} = -\frac{\kappa\sqrt{h(t)}}{A} + \frac{q_{in}(t)}{A}$$

$h(t)$ State (internal variable) of the system

Discrete-time water tank



$$h(t+1) = h(t) + \frac{\Delta t}{A} \cdot (q_{in}(t) - q_{out}(t))$$

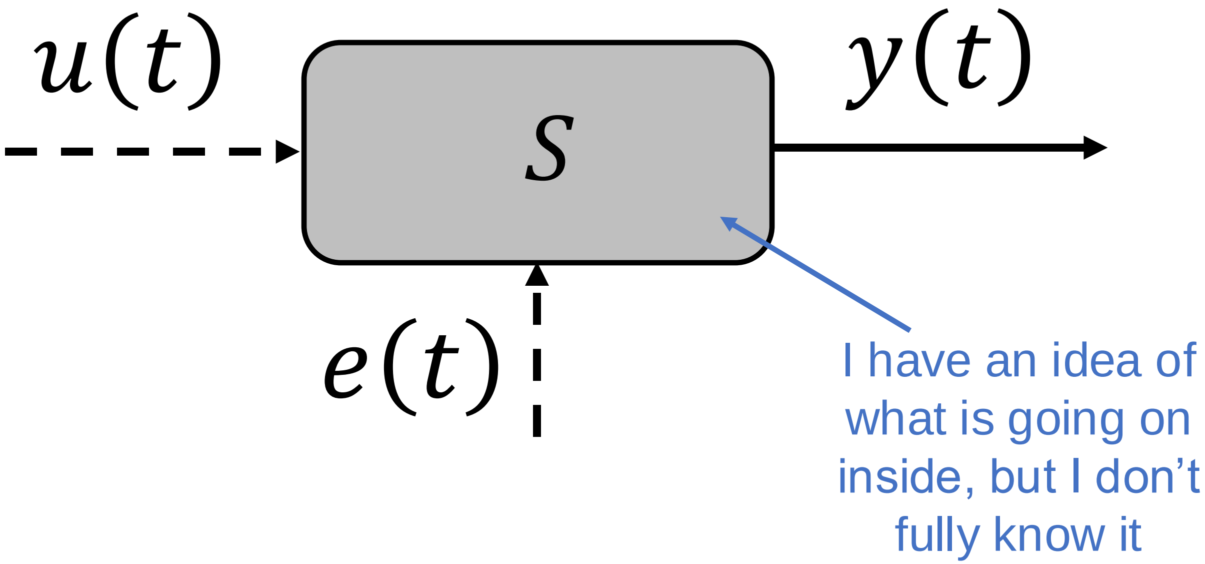
State: $x(t) = h(t)$
Input: $u(t) = q_{in}(t)$
Output: $y(t) = q_{out}(t)$

$$\begin{aligned} x(t+1) &= x(t) + \frac{\Delta t}{A}(u(t) - \kappa\sqrt{x(t)}) \\ y(t) &= \kappa\sqrt{x(t)} \end{aligned}$$

$$\begin{aligned} V(t+1) &= V(t) + q_{in}(t) \cdot \Delta t - q_{out}(t) \cdot \Delta t \\ A \cdot h(t+1) &= A \cdot h(t) + q_{in}(t) \cdot \Delta t - q_{out}(t) \cdot \Delta t \end{aligned}$$

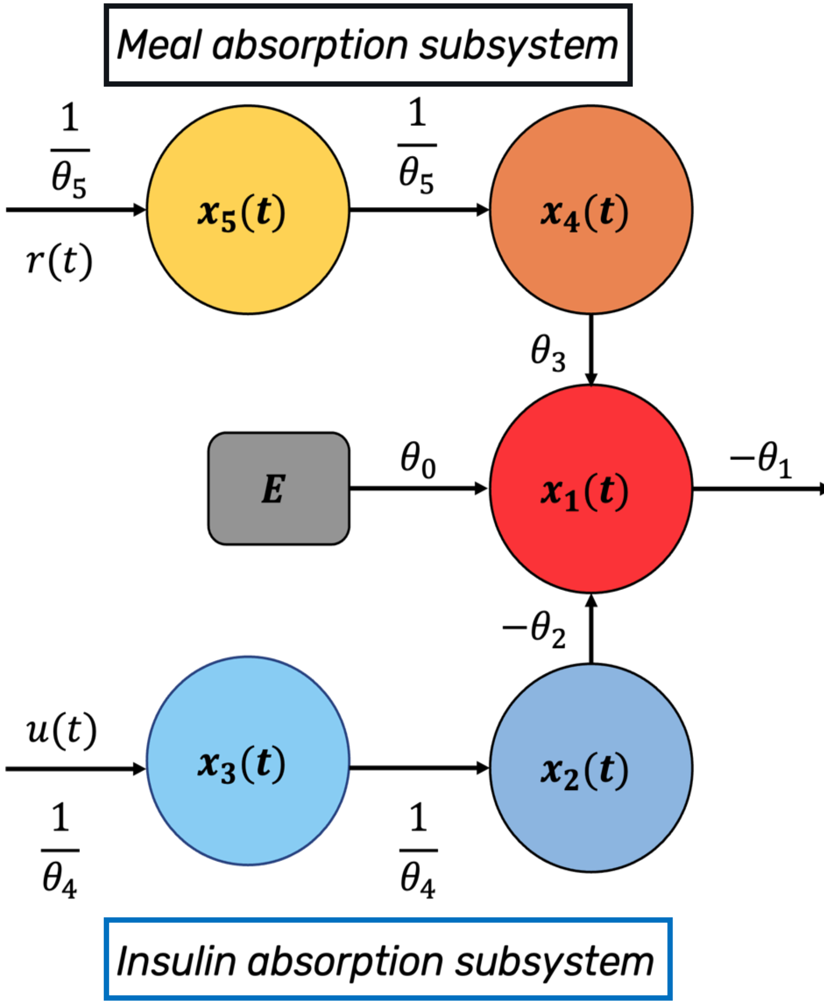
Grey-box model

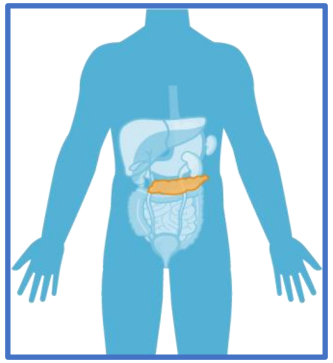
Grey-box model: either the structure of the model or the parameters are unknown



This happens when we know the behavior of the system is well described by a certain mathematical model, but we don't know it's parameters

Blood glucose regulation model





$$\frac{dQ_g(t)}{dt} = -\frac{1}{\theta_5} Q_g(t) + \frac{1}{\theta_5} Q_{sto}(t)$$

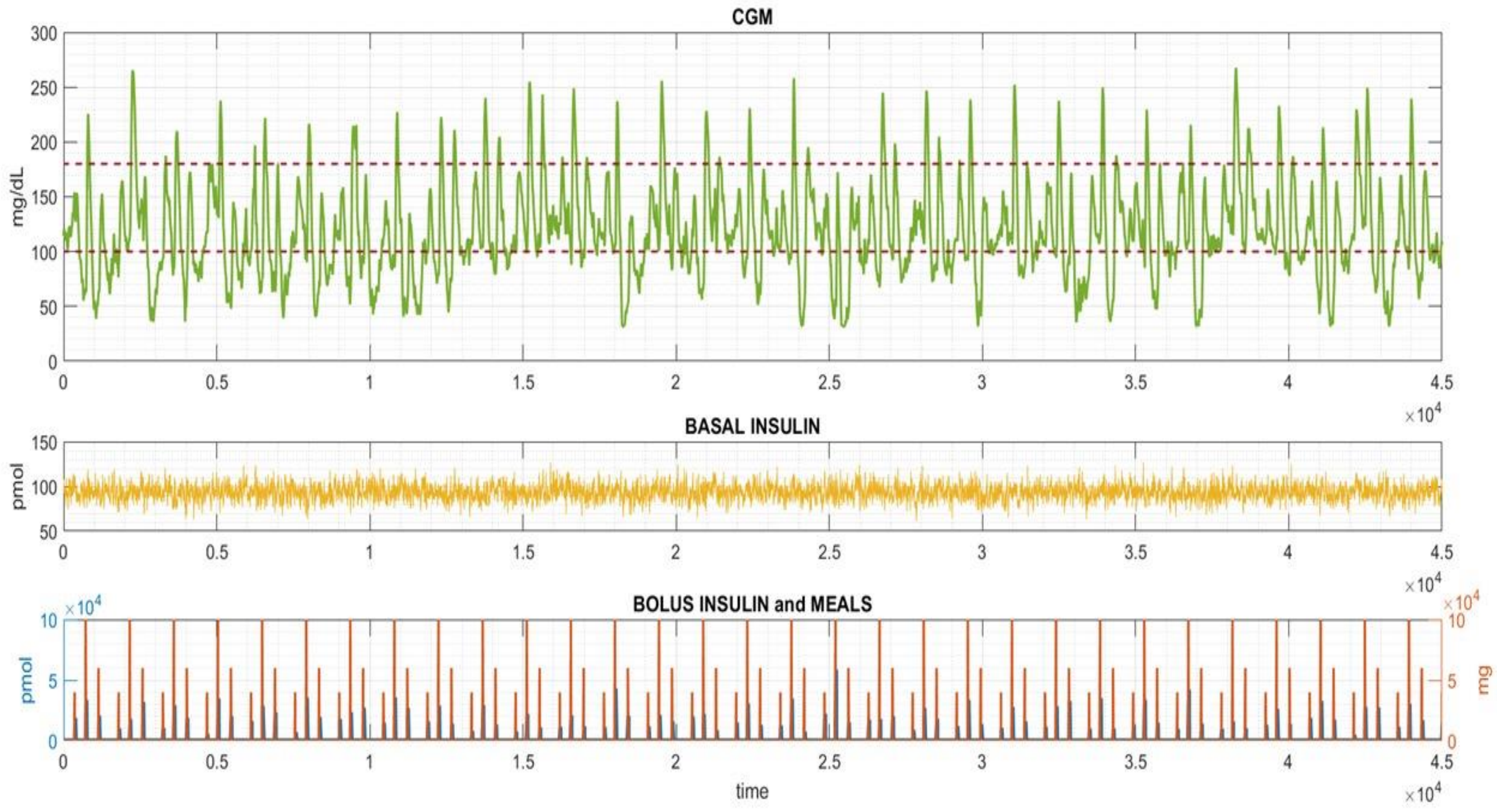
$$\frac{dQ_{sto}(t)}{dt} = -\frac{1}{\theta_5} Q_{sto}(t) + \frac{1}{\theta_5} r(t)$$

$$\frac{dG(t)}{dt} = \theta_0 - \theta_1 G(t) - \theta_2 Q_i(t) + \theta_3 Q_g(t)$$

$$\frac{dQ_i(t)}{dt} = -\frac{1}{\theta_4} Q_i(t) + \frac{1}{\theta_4} Q_{i_{sub}}(t)$$

$$\frac{dQ_{i_{sub}}(t)}{dt} = -\frac{1}{\theta_4} Q_{i_{sub}}(t) + \frac{1}{\theta_4} u(t)$$

Blood glucose regulation model - experiments

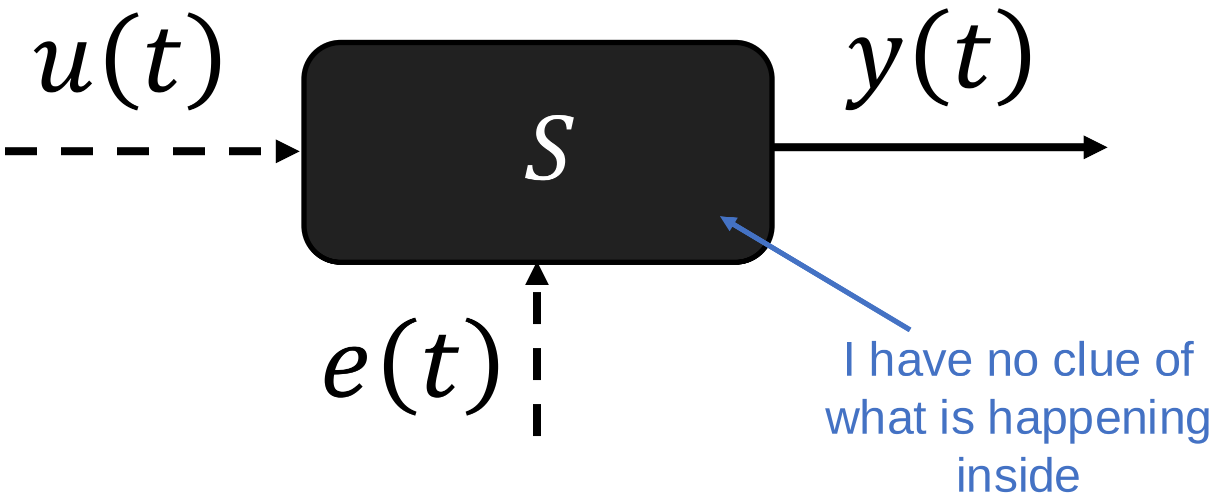


Running experiments we can collect data in different situations.

The model that we look for, needs to be a good approximation of these real experiments

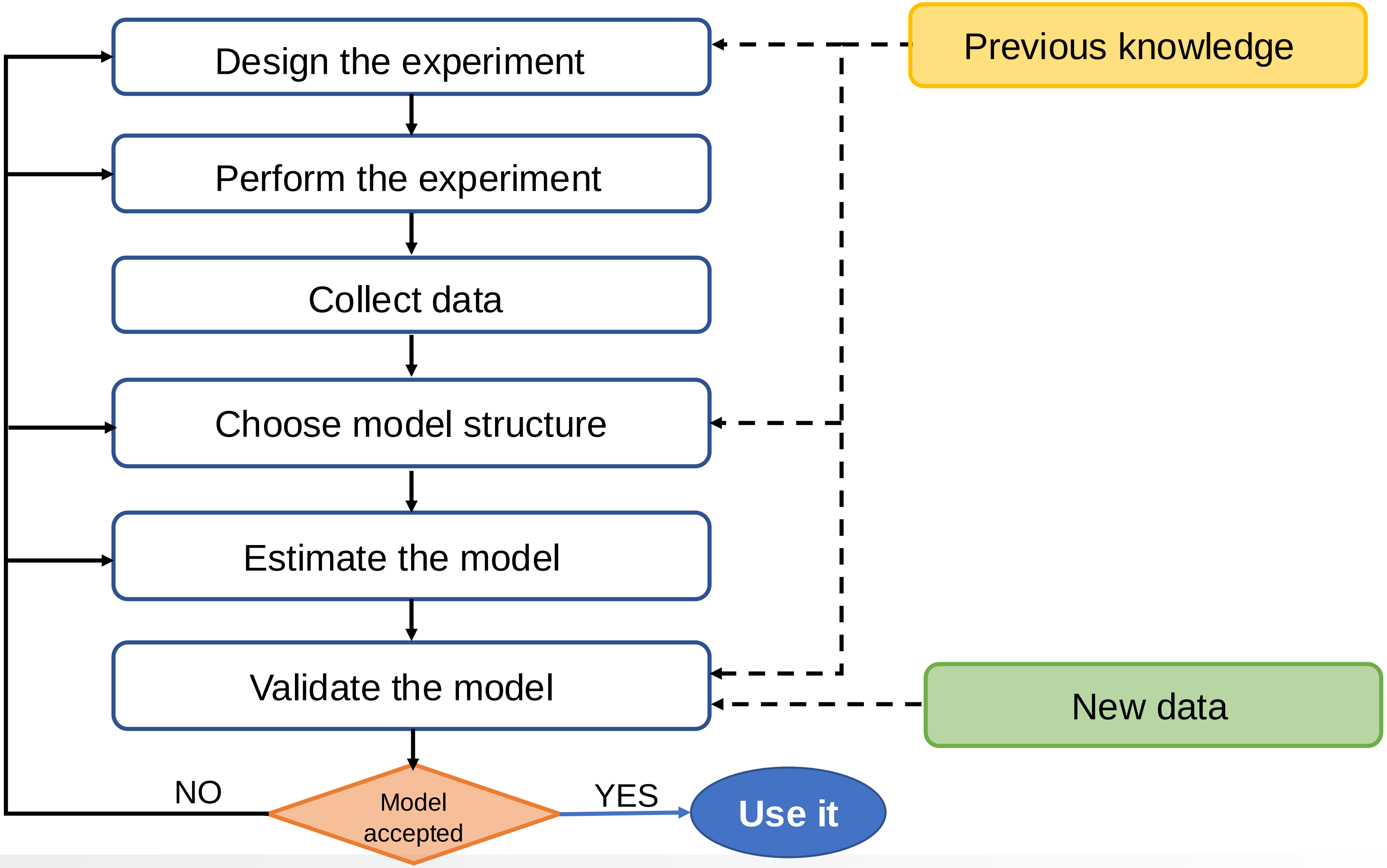
Black-box model

Black-box model: both the structure of the model and the parameters are unknown



In this case, the only thing we can use to construct the model are the **available data**.

System identification steps



System identification steps

1. Experiment design and **data collection**

2. Selection of a model class $\mathcal{M}(\boldsymbol{\theta}) = \{M(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$, where $\boldsymbol{\theta}$ is a **vector of parameters** (each different $\boldsymbol{\theta}$ corresponds to a different model). The model structure can be known a priori (white and grey box) or totally unknown (black box).

3. Choice of the identification criterion $J_N(\boldsymbol{\theta}) \geq 0$, that **measures the performance** of the model corresponding to $\boldsymbol{\theta}$ in describing available data. The best model is the one that **minimizes** this criterion $$\widehat{\boldsymbol{\theta}}_N = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, J_N(\boldsymbol{\theta})$$

System identification steps

4. **Minimization** of $J_N(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ (this minimization process will lead us to $\widehat{\boldsymbol{\theta}}_N$)

5. **Model validation.** Once the optimal model $M(\widehat{\boldsymbol{\theta}}_N)$ has been obtained, we verify whether this model is actually a good one. If it is not, the identification process must be repeated

Outline

1. Introduction to identification
2. **Experiment design and model selection**
3. Identification criterion
4. Model validation
5. A priori identifiabilty

Step 1: Experiment design and data collection

Issues when performing data collection:

1. Choice of the **data length** N
2. Design of the **input** $u(t)$ [for I/O systems]

We want input signals that “excite” all the system’s frequency components: that is, the system should be moved in almost all its output space

For system like pharmacokinetic models, we can choose different administration rates and observe how the behavior of the system changes due to different input rates.

Step 2: model class $\mathcal{M}(\boldsymbol{\theta}) = \{M(\boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$

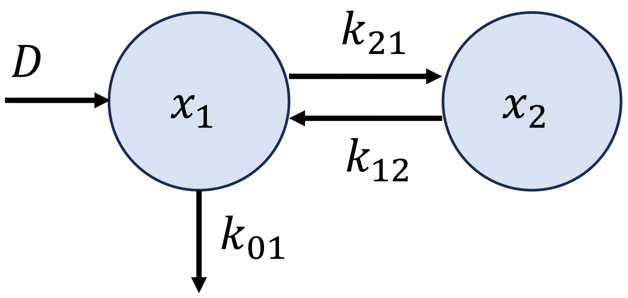
Many choices in general:

Discrete time	vs.	Continuous time
Linear	vs.	Nonlinear
Time invariant	vs.	Time variant

- ✓ Sometimes the model structure is already known from the physics/physiology of the system (ex. water tank).
- ✓ Sometimes we made up a structure based on some reasoning (compartmental models).
- ✓ Sometimes the structure is pure data-driven (ARX, NARX, Neural Networks).

Example

Given a compartmental model like the one of the previous example:



$$\dot{x}_1(t) = -(k_{01} + k_{21})x_1(t) + k_{12}x_2(t)$$

$$\dot{x}_2(t) = k_{21}x_1(t) - k_{12}x_2(t)$$

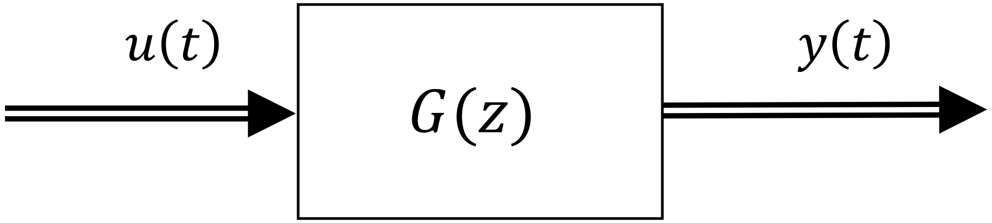
$$y(t) = \frac{x_1(t)}{V_1} \qquad D = 500$$

Then, the vector of parameters to be identified, $\boldsymbol{\theta}$ is given by:

$$\boldsymbol{\theta} = \{k_{01},\ k_{12},\ k_{21}, V_1\}$$

Example

Given a discrete time model in transfer function:



$$\mathcal{M}(\boldsymbol{\theta}): y(t) = \frac{b_0 + b_1 z^{-1}}{a_0 + a_1 z^{-1} + a_2 z^{-2}} u(t)$$

Then, the vector of parameters to be identified, $\boldsymbol{\theta}$ is given by:

$$\boldsymbol{\theta} = \{a_0, a_1, a_2, b_0, b_1\}$$

Step 3: choice of the identification criterion

$J_N(\boldsymbol{\theta})$ must measure the capability of model $M(\boldsymbol{\theta})$ in describing the collected data

$$\{u(1), \dots, u(N), y(1), \dots, y(N)\}$$

Predictive approach: generate predictors $\hat{y}(t, \boldsymbol{\theta})$ from models and evaluate the model capability of **predicting the system behavior**

- Predictor must be fed with the same inputs that we used to obtain the output date from the experiment, so we can evaluate the performance of the model coherently

- The best model is the one with the best predictive performance

Outline

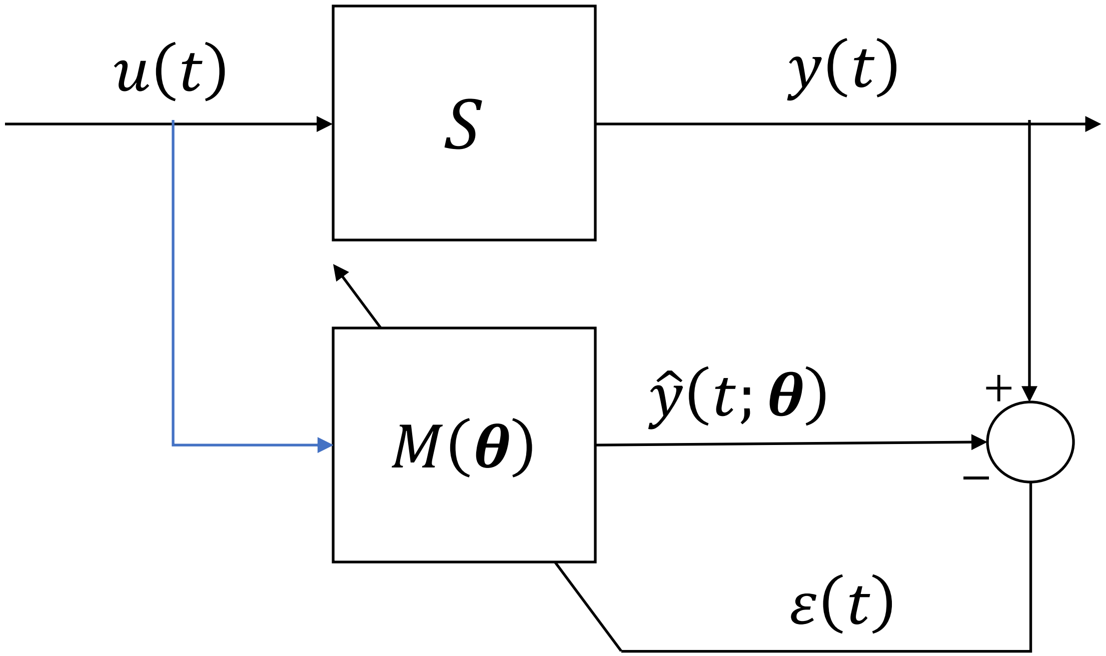
1. Introduction to identification
2. Experiment design and model selection
3. **Identification criterion**
4. Model validation
5. A priori identifiabilty

Output error identification scheme

Identification criterion

$$J_N(\boldsymbol{\theta}) = \frac{1}{N}\sum_{k=1}^{N}\big(y(k) - \hat{y}(t;\boldsymbol{\theta})\big)^2$$

We want to minimize **the prediction error,** that is the difference between real output and the one predicted by the model



Best model

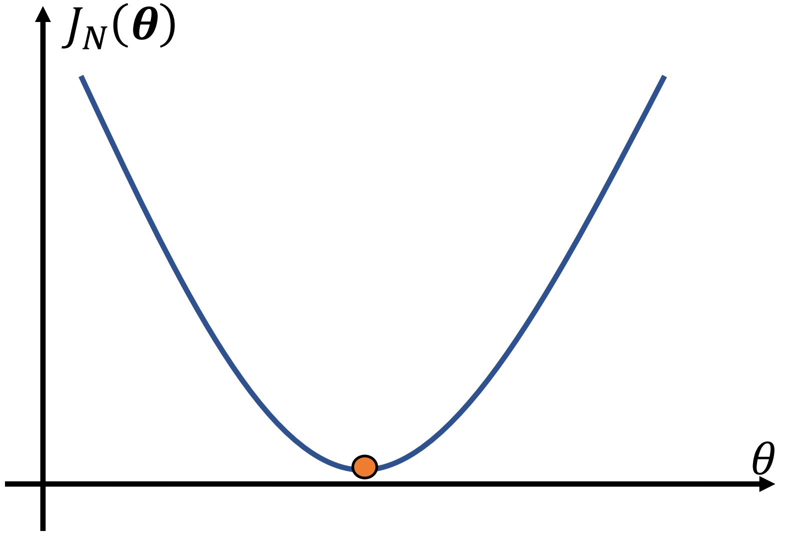
$$\widehat{\boldsymbol{\theta}}_N = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, J_N(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\big(y(k) - \hat{y}(t;\boldsymbol{\theta})\big)^2$$

i.e. the best model is the one minimizing the **prediction error**

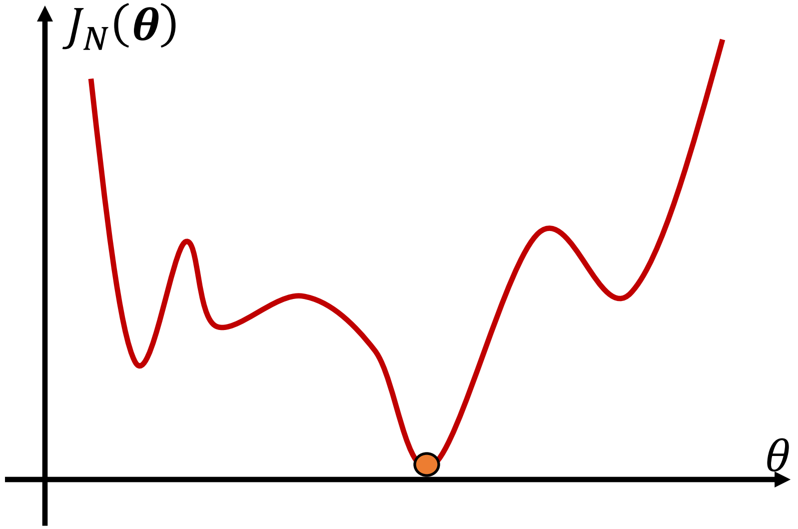
Step 4: Minimization of $J_N(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$

Two relavant cases, depending on the model structure:

- $J_N(\boldsymbol{\theta})$ is **quadratic ➔ UNIQUE GLOBAL MINIMUM** (also in closed form)

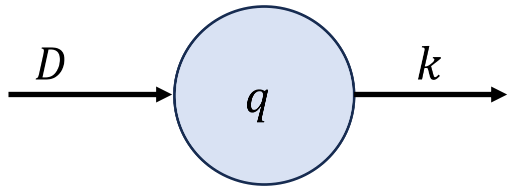


- $J_N(\boldsymbol{\theta})$ is **not quadratic ➔ LOCAL MINIMA** (iterative optimization algorithms)
  - Newton and Quasi-Newton methods



Identification of compartmental models

Let's analyze the system



$$\dot{q}(t) = -kq(t) + D\delta(t) \qquad q(0) = 0$$

$$q(t) = De^{-kt}$$

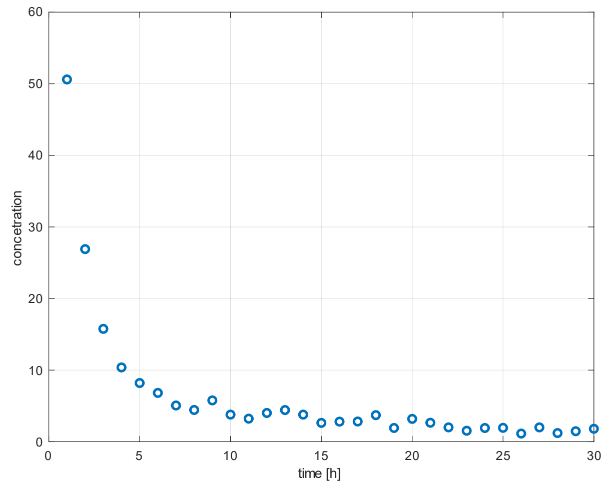
$$c(t) = \frac{q(t)}{V}$$

Output of the model

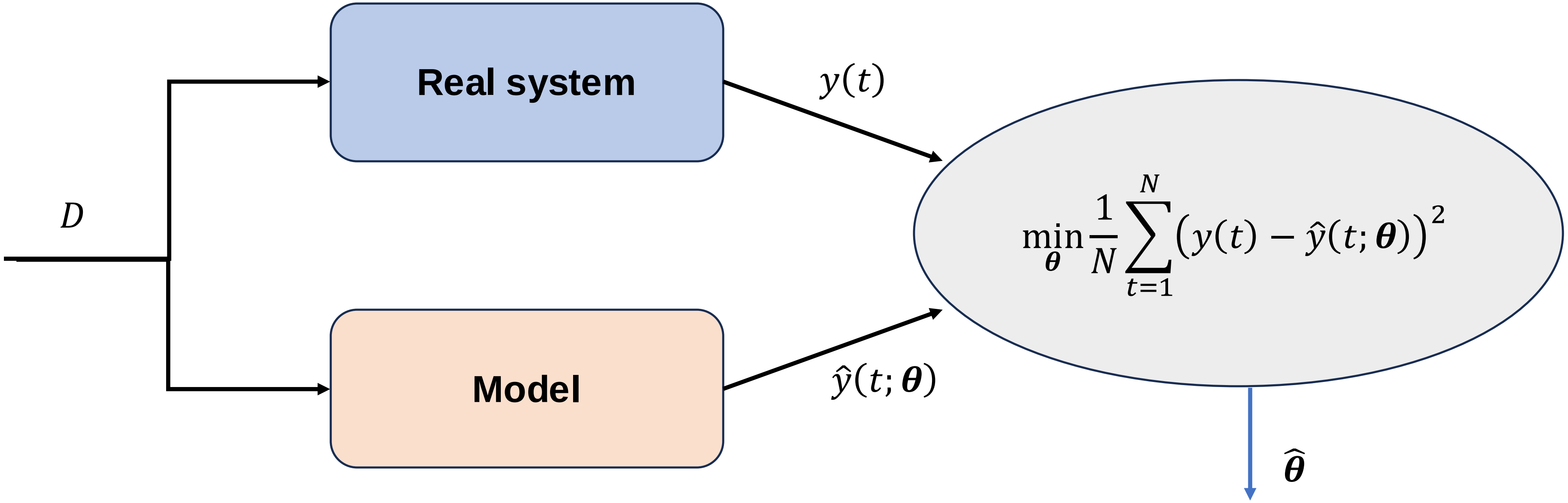
Identifying this model means to determine the values of the parameters

$$\boldsymbol{\theta} = \{k, V\}$$

We run experiments with different values of D to get output measurements like this



Identification of compartmental models



Using softwares like Matlab, we cast an optimization problem in which we try to solve the previous problem by minimizing the output error

Identification of compartmental models

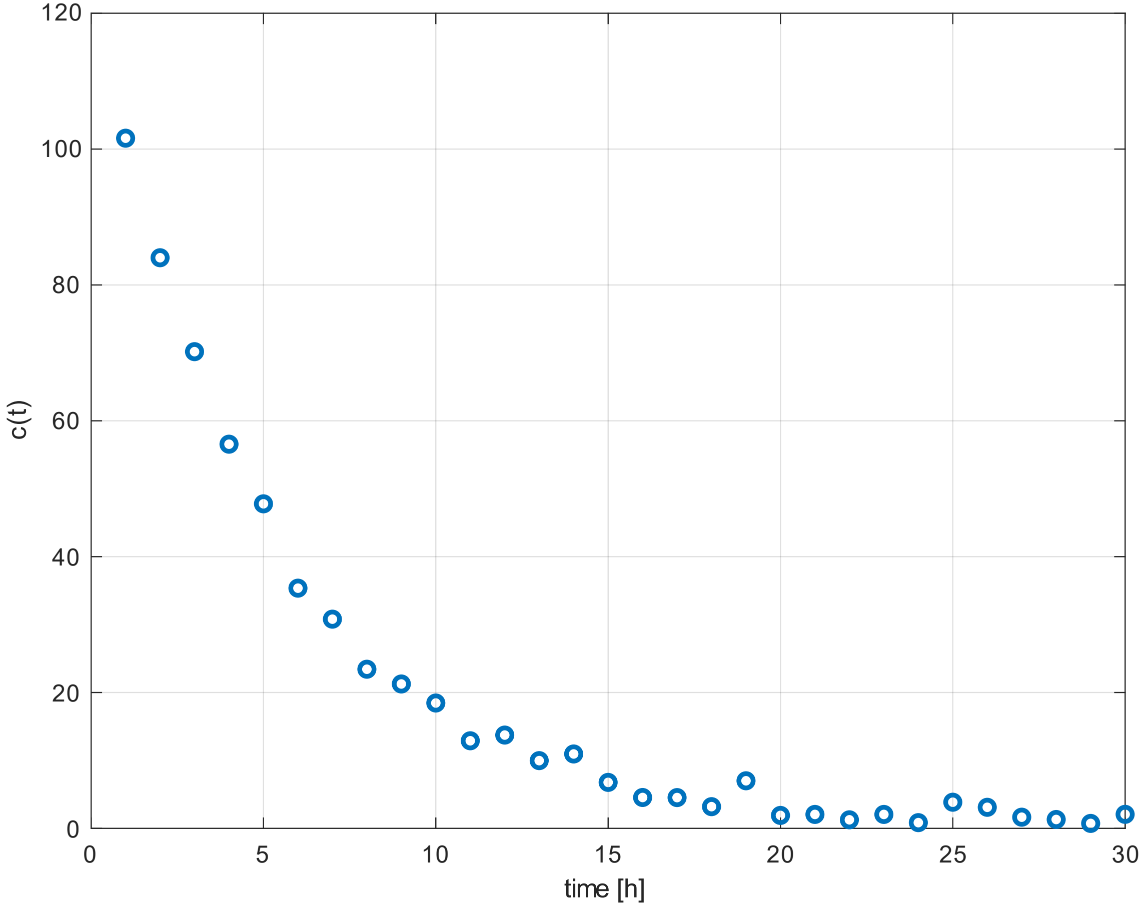
Using softwares like Matlab, we cast an optimization problem to solve the identification by minimizing the output error

```
%------------------------------------%
%% Optimization problem (resolution)
%------------------------------------%

theta0=0.1*ones(2,1);
N=Tmax;

t_min=zeros(2,1);
t_max=[0.5;15];

[theta_o,FVAL]=fmincon(@(theta)cost_function(theta,x0,yd,N),theta0,...
                        [],[],[],[],t_min,t_max);

k=theta_o(1);
V=theta_o(2);

Am=[-k];

ode=@(t,x) [Am*x];

[tm,xs]=ode45(ode,[1,Tmax],x0);

ym=xs(:,1)/V;

figure
scatter(td,yd)
hold on
plot(tm,ym)
```

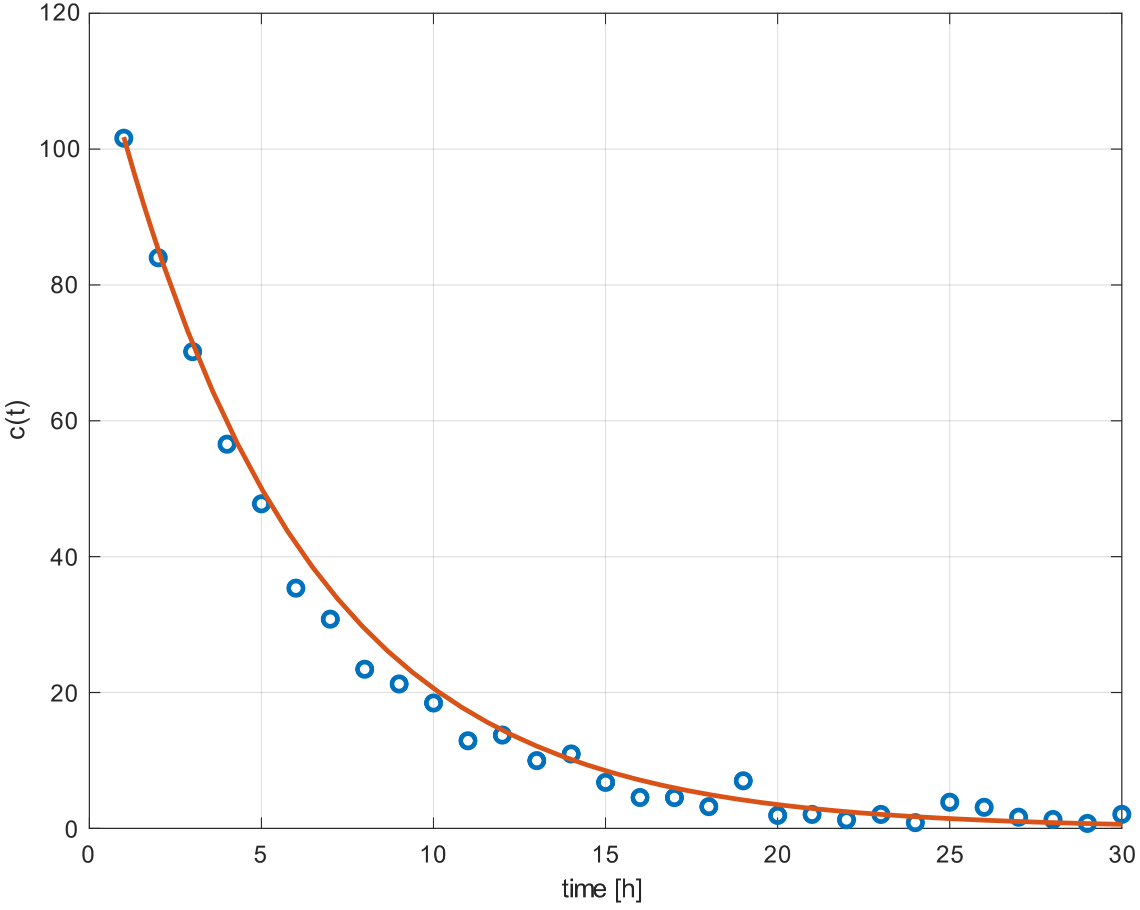
```
function J=cost_function(theta,x0,yd,N)

k=theta(1);
V=theta(2);

Am=[-k];

Ts=1;

J=0;
x(:,1)=x0;

for k=1:N
   x(:,k+1)=x(:,k)+Am*x(:,k)*Ts;
   tm(k)=k;
   ym(k)=(x(1,k)/V);
   J=J+(yd(k)-ym(k))^2;
end

J=J/N;

end
```

Identification of compartmental models

Measured output



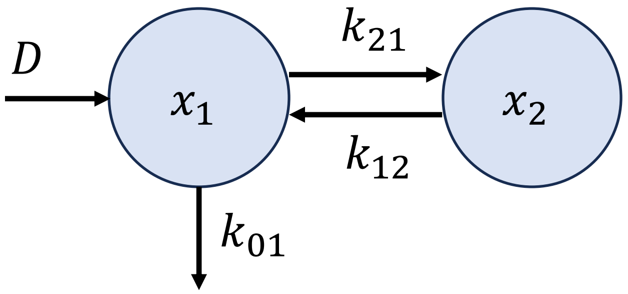
Output of the identified model



$$\boldsymbol{\theta} = \{k, V\} = \{0.18; 4.9\}$$

Identification of compartmental models

Let’s analyze a second order system



$$\dot{x_1}(t) = -(k_{01} + k_{21})x_1(t) + k_{12}x_2(t) + D\delta(t)$$

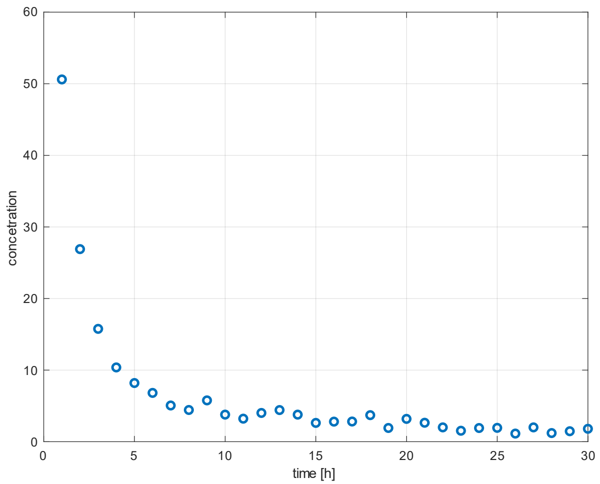
$$\dot{x_2}(t) = k_{21}x_1(t) - k_{12}x_2(t)$$

$$y(t) = \frac{x_1(t)}{V_1}$$ *Output of the model*

Identifying this model means to determine the values of the parameters

$$\boldsymbol{\theta} = \{k_{01},\ k_{12},\ k_{21}, V_1\}$$

We run experiments with different values of D to get output measurements like this



Identification of compartmental models

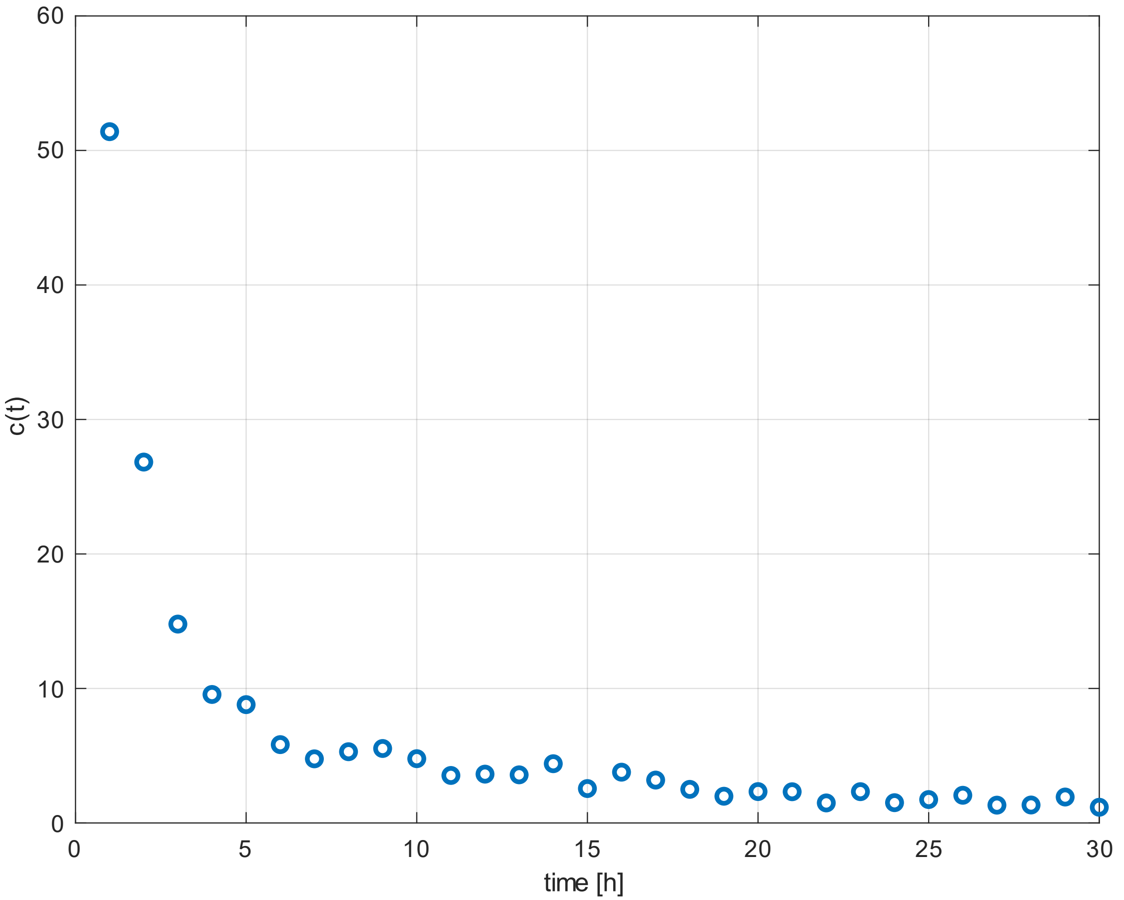
Using softwares like Matlab, we cast an optimization problem to solve the identification by minimizing the output error

```
%-------------------------------------%
%% Optimization problem (resolution)
%-------------------------------------%

theta0=0.1*ones(4,1);
N=Tmax;

t_min=zeros(4,1);
t_max=[0.3;0.3;0.3;15];

[theta_o,FVAL]=fmincon(@(theta)cost_function(theta,x0,yd,N),theta0,...
                       [],[],[],[],t_min,t_max);

k01=theta_o(1);
k21=theta_o(2);
k12=theta_o(3);
V=theta_o(4);

Am=[-k01-k21 k12; k21 -k12];

ode=@(t,x)[Am*x];

[tm,xs]=ode45(ode,[1,Tmax],x0);

ym=xs(:,1)/V;

figure
scatter(td,yd)
hold on
plot(tm,ym)
```

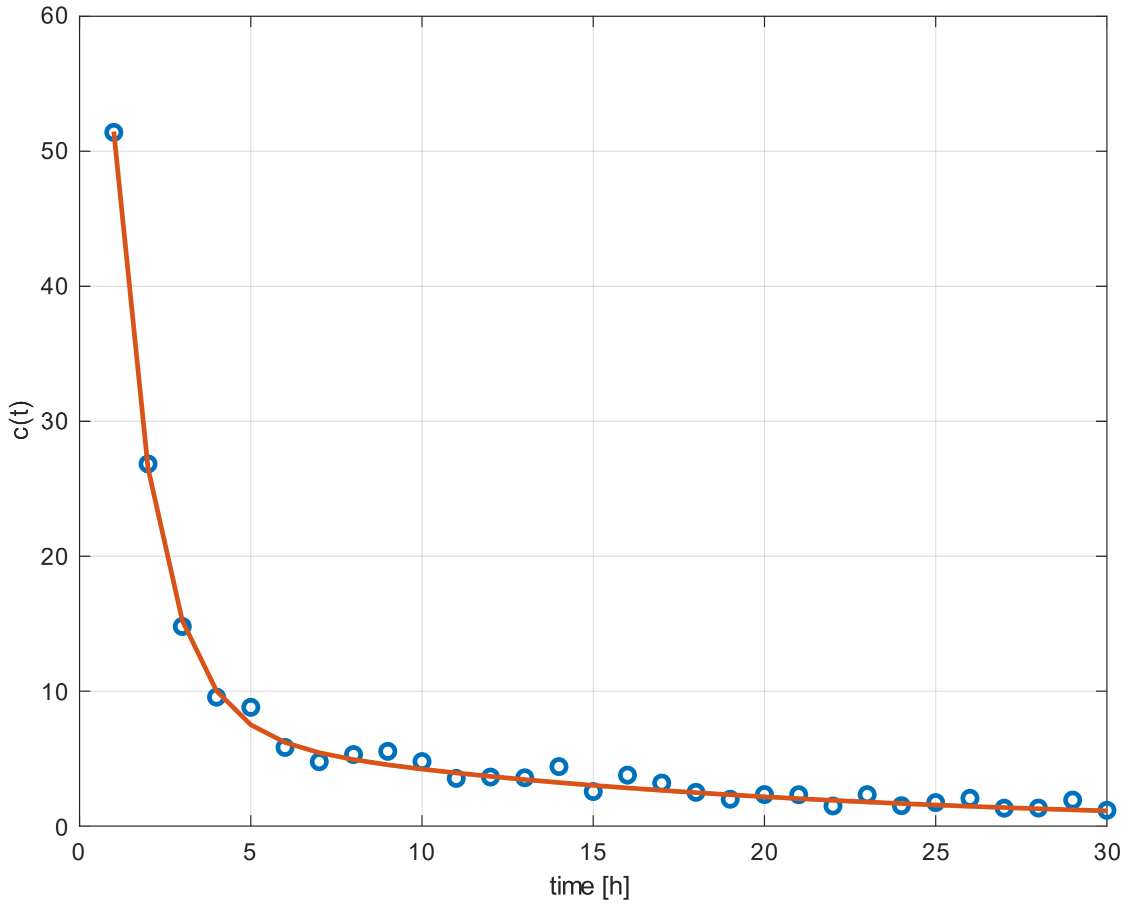
```
function J=cost_function(theta,x0,yd,N)

k01=theta(1);
k21=theta(2);
k12=theta(3);
V=theta(4);

Am=[-k01-k21 k12; k21 -k12];

Ts=1;

J=0;
x(:,1)=x0;

for k=1:N
   x(:,k+1)=x(:,k)+Am*x(:,k)*Ts;
   tm(k)=k;
   ym(k)=(x(1,k)/V);
   J=J+(yd(k)-ym(k))^2;
end

J=J/N;

end
```

Identification of compartmental models

Measured output



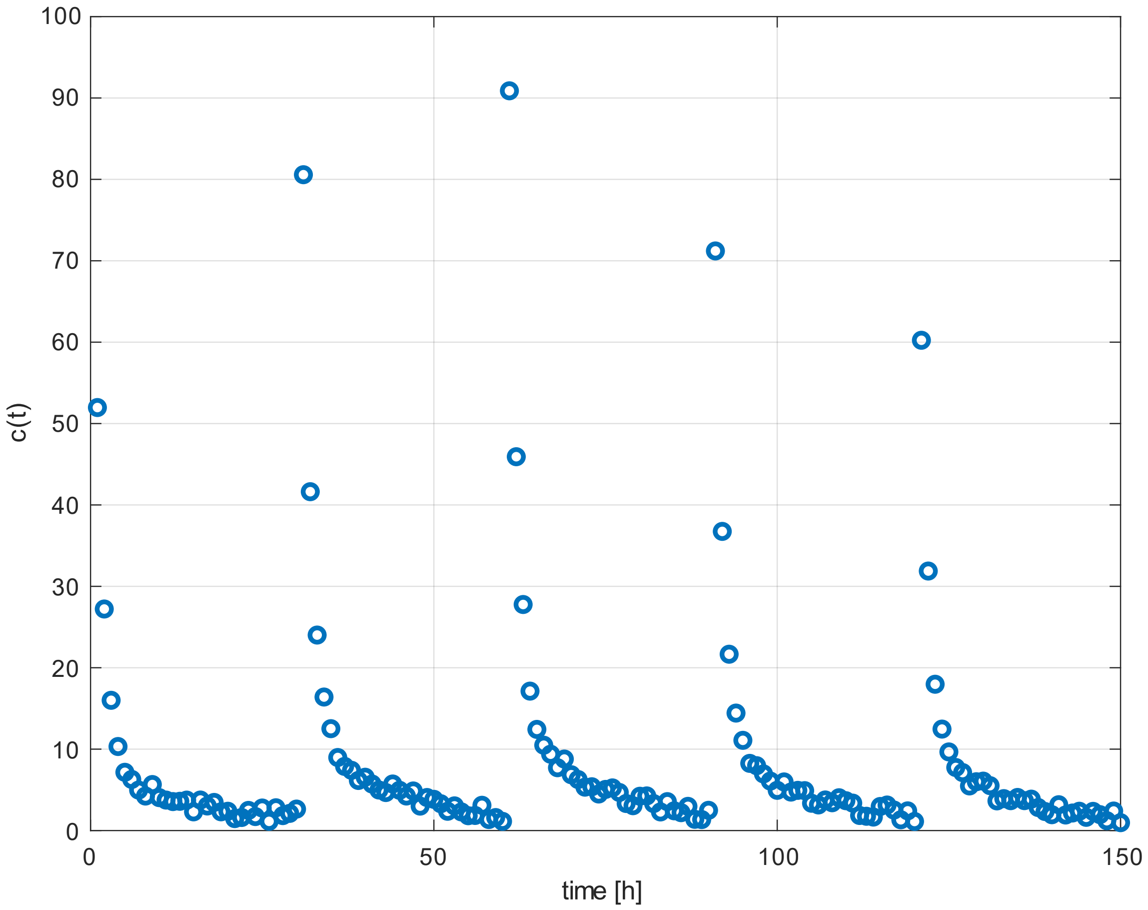
Output of the identified model



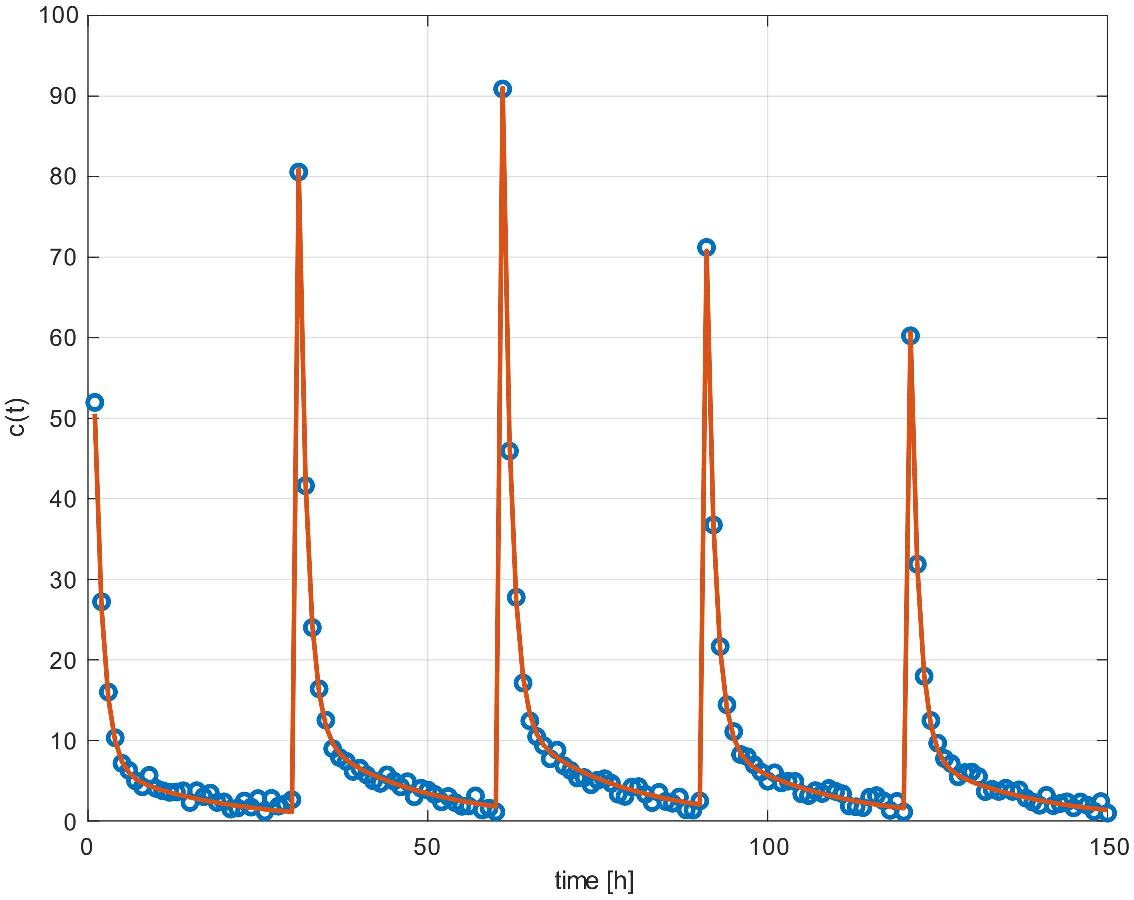
$$\boldsymbol{\theta} = \{k_{01},\ k_{12},\ k_{21}, V_1\} = \{0.24; 0.24; 0.12; 9.65\}$$

Identification of compartmental models

Measured output



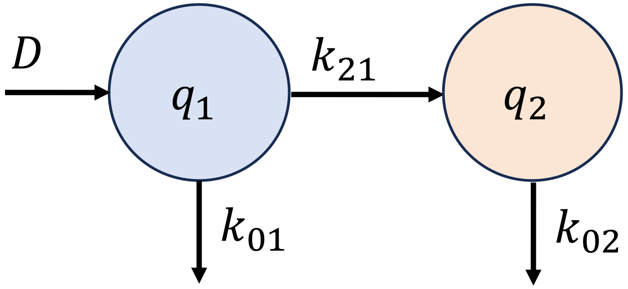
Output of the identified model



$$\boldsymbol{\theta} = \{k_{01},\ k_{12},\ k_{21}, V_1\} = \{0.26; 0.22; 0.13; 9.89\}$$

Identification of compartmental models

Let's analyze a system with absorption



$$\dot{q}_1(t) = -(k_{01} + k_{21})q_1(t) + D\delta(t)$$

$$\dot{q}_2(t) = k_{21}q_1(t) - k_{02}q_2(t)$$

$$F = \frac{k_{21}}{k_{01} + k_{21}}$$

$$y(t) = \frac{q_2(t)}{V_2}$$

Output of the model

Identifying this model means to determine the values of the parameters

$$\boldsymbol{\theta} = \{k_{01},\ k_{21},\ k_{02}, V_2\}$$

As in the previous cases, we can run experiments with different values of D to get output measurements and then compute the best set of parameters by minimizing the output error

Identification of compartmental models

This case is interesting since we can set the constraint on F

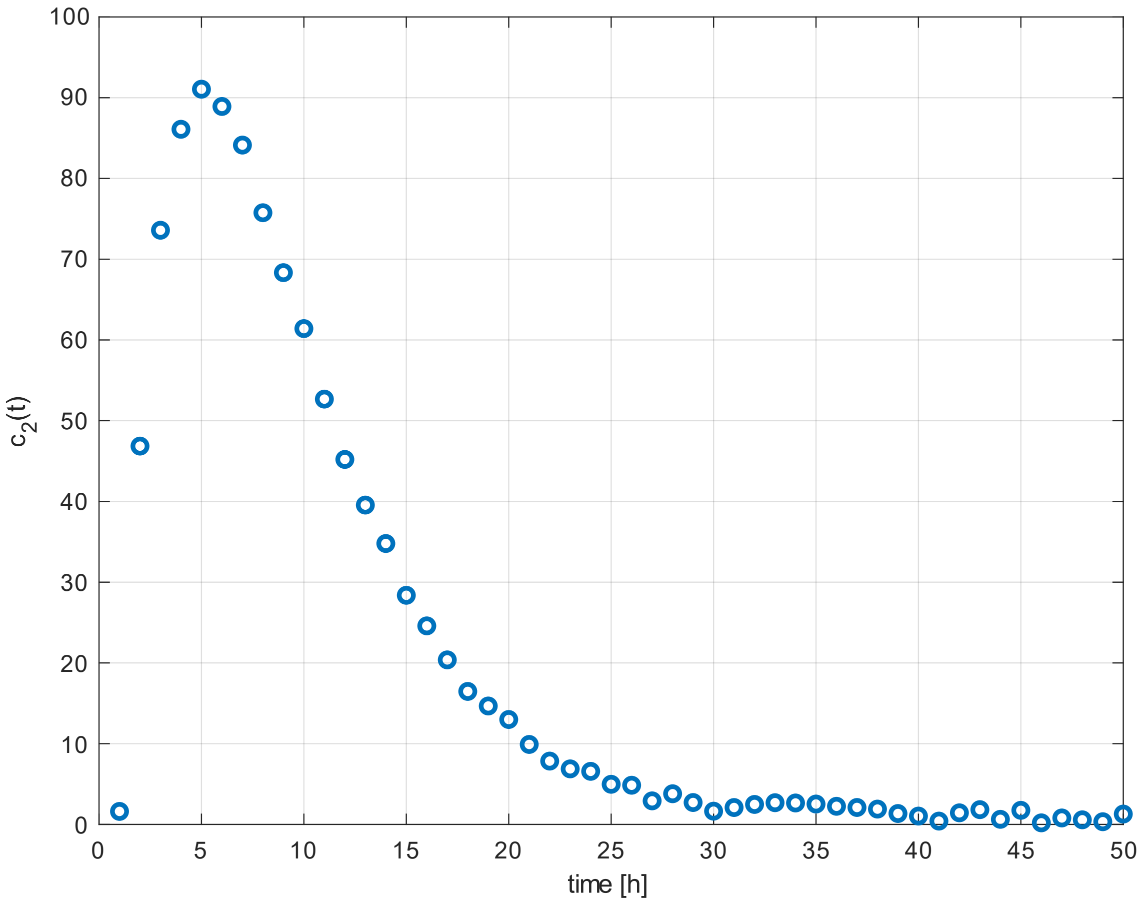
```
%-------------------------------------%
%% Optimization problem (resolution)
%-------------------------------------%

theta0=0.1*ones(4,1);
N=Tmax;

t_min=[zeros(4,1)];
t_max=[0.3;0.3;0.3;10];

[theta_o,FVAL]=fmincon(@(theta)cost_function(theta,x0,yd,N),theta0,...
                       [],[],[-F (1-F) 0 0],[0],t_min,t_max);

k01=theta_o(1);
k21=theta_o(2);
k02=theta_o(3);
V=theta_o(4);

Am=[-k01-k21 0; k21 -k02];

ode=@(t,x)[Am*x];

[tm,xs]=ode45(ode,[1,Tmax],x0);

ym=xs(:,2)/V;

figure
scatter(td,yd)
hold on
plot(tm,ym)
```

$$F = \frac{k_{21}}{k_{01} + k_{21}}$$

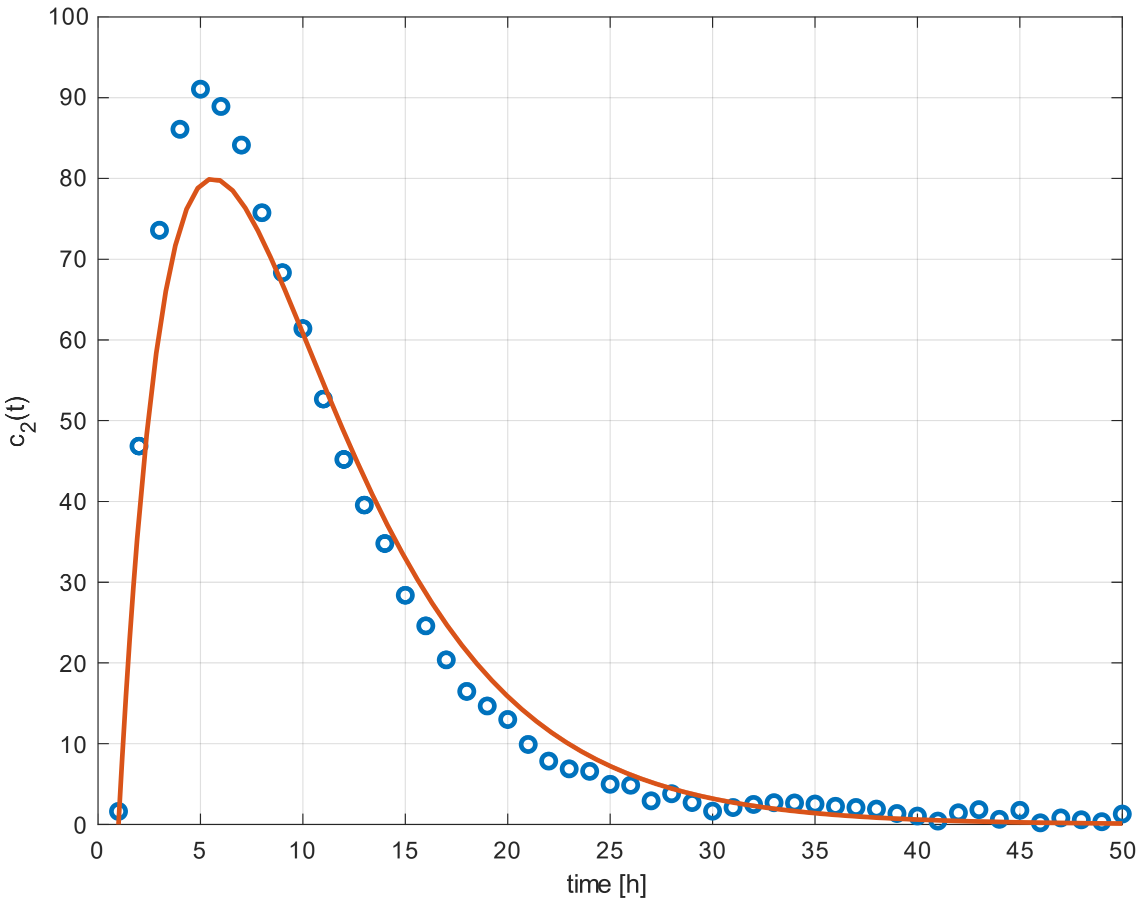
```
function J=cost_function(theta,x0,yd,N)

k01=theta(1);
k21=theta(2);
k02=theta(3);
V=theta(4);

Am=[-k01-k21 0; k21 -k02];

Ts=1;

J=0;
x(:,1)=x0;

for k=1:N
    x(:,k+1)=x(:,k)+Am*x(:,k)*Ts;
    tm(k)=k;
    ym(k)=(x(2,k)/V);
    J=J+(yd(k)-ym(k))^2;
end

J=J/N;

end
```

Identification of compartmental models

Measured output



Output of the identified model



$$\boldsymbol{\theta} = \{k_{01},\ k_{21},\ k_{02}, V_2\} = \{0.05; 0.22; 0.17; 2.27\}$$

Identification of linear discrete-time models

Consider a generic linear discrete time model:

$$\mathcal{M}(\boldsymbol{\theta}):\ \ y(t) = \frac{b_0 + b_1 z^{-1} + \cdots + b_p z^{-p}}{1 - a_1 z^{-1} - \cdots - a_m z^{-m}} u(t) = \frac{B(z,\boldsymbol{\theta})}{A(z,\boldsymbol{\theta})} u(t)$$

$$A(z) = 1 - a_1 z^{-1} - \cdots - a_m z^{-m}$$

$$B(z) = b_0 + b_1 z^{-1} + \cdots + b_p z^{-p}$$

$$\boldsymbol{\theta} = [a_1 \quad \cdots \quad a_m \quad b_0 \quad b_1 \quad \cdots \quad b_p\,]^\top$$

Output predicted by the model: $\hat{y}(t;\boldsymbol{\theta}) = G(z,\boldsymbol{\theta}) u(t)$

Recursive representation:

$$\hat{y}(t;\boldsymbol{\theta}) = a_1 y(t-1) + \cdots + a_m y(t-m) + b_0 u(t) + b_1 u(t-1) + \cdots + b_p u(t-p)$$

Identification of linear discrete-time models

Write the model in vector form:

$$\hat{y}(t;\boldsymbol{\theta}) = a_1 y(t-1) + \cdots + a_m y(t-m) + b_0 u(t) + b_1 u(t-1) + \cdots + b_p u(t-p)$$

$$= \underset{1\times d}{\boldsymbol{\varphi}(t)^\top} \underset{d\times 1}{\boldsymbol{\theta}}$$

- $\boldsymbol{\varphi}(t) = [y(t-1) \quad \cdots \quad y(t-m) \quad u(t) \quad u(t-1) \quad \cdots \quad u(t-p)]^\top$
- $\boldsymbol{\theta} = [a_1 \quad \cdots \quad a_m \quad b_0 \quad b_1 \quad \cdots \quad b_p]^\top$

$$\hat{y}(t;\boldsymbol{\theta}) = \boldsymbol{\varphi}(t)^\top \boldsymbol{\theta}$$

$$y(t) - \hat{y}(t;\boldsymbol{\theta}) = y(t) - \boldsymbol{\varphi}(t)^\top \boldsymbol{\theta}$$

Identification of linear discrete-time models

Identification criterion

$$J_N(\boldsymbol{\theta}) = \frac{1}{N}\sum_{t=1}^{N}\left(y(t) - \hat{y}(t;\boldsymbol{\theta})\right)^2 = \frac{1}{N}\sum_{t=1}^{N}\left(y(t) - \boldsymbol{\varphi}(t)^\top\boldsymbol{\theta}\right)^2$$

Since $\boldsymbol{\varphi}(t)^\top\boldsymbol{\theta}$ is **linear** in $\boldsymbol{\theta}$, the cost function $J_N(\boldsymbol{\theta})$ is **quadratic** ➔ **minimum can be explicitly computed.** We can obtain the solution by imposing:

$$-\frac{2}{N}\sum_{t=1}^{N}\left(y(t) - \boldsymbol{\varphi}(t)^\top\boldsymbol{\theta}\right)\boldsymbol{\varphi}(t)^\top = \mathbf{0}$$

Identification of linear discrete-time models

How do we get to this?

- Optimization theory gives us the condition to find the minimun:

$$\frac{dJ_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = 0$$

The derivative vector must be null, which is the condition to find stationary points

$$\frac{d^2J_N(\boldsymbol{\theta})}{d^2\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \geq 0$$

The Hessian matrix must be positive semi-definite, condition for spotting out minima

Identification of linear discrete-time models

The derivative vector is a column vector given by:

$$\frac{dJ_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}} = \frac{d}{d\boldsymbol{\theta}}\left[\frac{1}{N}\sum_{t=1}^{N}\left(y(t)-\boldsymbol{\varphi}(t)^\top\boldsymbol{\theta}\right)^2\right] = \frac{1}{N}\sum_{t=1}^{N}\frac{d}{d\boldsymbol{\theta}}\left(y(t)-\boldsymbol{\varphi}(t)^\top\boldsymbol{\theta}\right)^2$$

$$= \frac{1}{N}\sum_{t=1}^{N}2\left(y(t)-\boldsymbol{\varphi}(t)^\top\boldsymbol{\theta}\right)\left(-\boldsymbol{\varphi}(t)^\top\right) = -\frac{2}{N}\sum_{t=1}^{N}\left(y(t)-\boldsymbol{\varphi}(t)^\top\boldsymbol{\theta}\right)\left(\boldsymbol{\varphi}(t)^\top\right)$$

And it should be null in the minimum. Then:

$$-\frac{2}{N}\sum_{t=1}^{N}\left(y(t)-\boldsymbol{\varphi}(t)^\top\boldsymbol{\theta}\right)\boldsymbol{\varphi}(t)^\top = \mathbf{0}$$

Identification of linear discrete-time models

Transposing and separating terms, one obtains the **normal equations:**

$$\widehat{\boldsymbol{\theta}} = \left(\sum_{t=1}^{N} \boldsymbol{\varphi}(t)\boldsymbol{\varphi}(t)^\top \right)^{-1} \cdot \sum_{t=1}^{N} \boldsymbol{\varphi}(t)y(t)$$

This is the **least squares method.**

The solution to the normal equation are always **minimum points**, since the Hessian is always positive semi-definite. In fact:

$$\frac{d^2 J_N(\boldsymbol{\theta})}{d^2\boldsymbol{\theta}} = \frac{d}{d\boldsymbol{\theta}} \frac{dJ_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}} = \frac{2}{N} \sum_{t=1}^{N} \boldsymbol{\varphi}(t)\,\boldsymbol{\varphi}(t)^\top$$

This matrix is always positive semi-definite

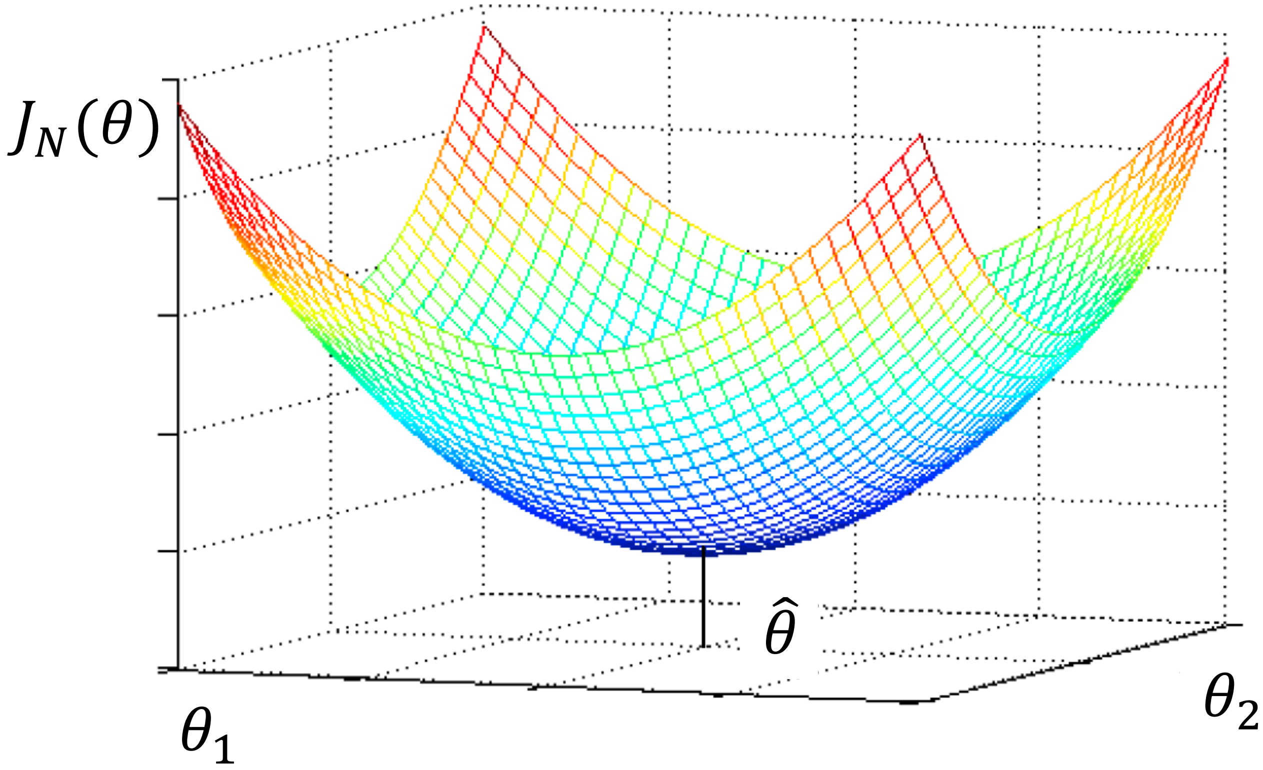
(A matrix M is positive semi-definite if $x^T M x \geq 0 \ \forall x \neq 0$*)*

Identification of linear discrete-time models

There are 2 possibilities:

- **Case 1**: $\dfrac{d^2 J_N(\boldsymbol{\theta})}{d^2\boldsymbol{\theta}}$ is non singular → then it's invertible

Unique minimizer of $J_N(\boldsymbol{\theta})$



UNIVERSITÀ DEGLI STUDI DI BERGAMO
Dipartimento di Ingegneria Gestionale, dell'Informazione e della Produzione

Identification of linear discrete-time models

There are 2 possibilities:

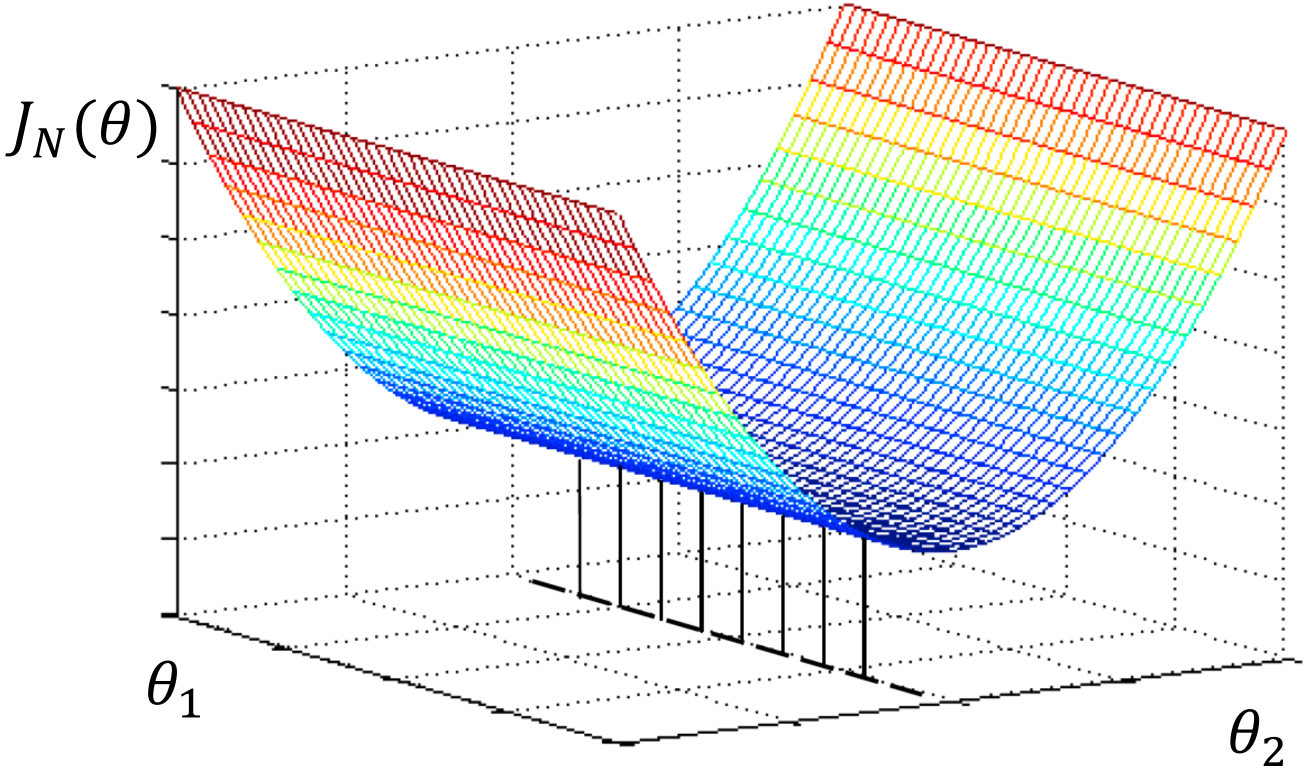
- **Case 2**: $\dfrac{d^2 J_N(\boldsymbol{\theta})}{d^2\boldsymbol{\theta}}$ is singular → then it's not invertible!

Infinite minimum points of $J_N(\boldsymbol{\theta})$

All solution are equivalent.

WARNING: this result means that:

1. Data are not good enough
2. The chosen model is too complex



Example

Suppose that $N = 10$ data from an I/O system are available

$$\{y(1), y(2), \dots, y(N)\}$$

$$\{u(1), u(2), \dots, u(N)\}$$

Estimate an model $\quad y(t) = \dfrac{b}{1 + az^{-1}} u(t-1)$

$$y(t) = -ay(t-1) + bu(t-1)$$

⬇

$$\hat{y}(t; \boldsymbol{\theta}) = -ay(t-1) + bu(t-1)$$

Example

$$\hat{y}(t;\boldsymbol{\theta}) = -ay(t-1) + bu(t-1)$$

$$J_{10}(a,b) = \frac{1}{9}\sum_{t=2}^{10}\left(y(t) - \hat{y}(t;\boldsymbol{\theta})\right)^2 = \frac{1}{9}\sum_{t=2}^{10}\left(y(t) + ay(t-1) - bu(t-1)\right)^2$$

$$\underset{N\times d}{\Phi} = \begin{bmatrix} y(1) & u(1) \\ y(2) & u(2) \\ \vdots & \vdots \\ y(9) & u(9) \end{bmatrix} \qquad \underset{N\times 1}{Y} = \begin{bmatrix} y(2) \\ y(3) \\ \vdots \\ y(10) \end{bmatrix} \qquad \underset{d\times 1}{\boldsymbol{\theta}} = \begin{bmatrix} a \\ b \end{bmatrix}$$

$$\hat{\boldsymbol{\theta}}_N = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \|Y - \Phi\boldsymbol{\theta}\|_2^2$$

$$\boxed{\hat{\boldsymbol{\theta}}_N = (\Phi^\top\Phi)^{-1}\Phi^\top Y}$$

Example

$$\hat{y}(t;\boldsymbol{\theta}) = -ay(t-1) + bu(t-1)$$

	1	2	3	4	5	6	7	8	9	10
u(t)	0.5	1	1.3	1.7	2.1	1.6	0.5	0.8	1.4	2
y(t)	0.17	0.23	0.76	0.91	1.53	1.14	0.83	0.95	1.2	1.5

$N \times d$

$$\Phi = \begin{bmatrix} 0.17 & 0.5 \\ 0.23 & 1 \\ 0.76 & 1.3 \\ 0.91 & 1.7 \\ 1.53 & 2.1 \\ 1.14 & 1.6 \\ 0.83 & 0.5 \\ 0.95 & 0.8 \\ 1.2 & 1.4 \end{bmatrix}$$

$N \times 1$

$$Y = \begin{bmatrix} 0.23 \\ 0.76 \\ 0.91 \\ 1.53 \\ 1.14 \\ 0.83 \\ 0.95 \\ 1.2 \\ 1.5 \end{bmatrix}$$

$$\Phi^\top \Phi = \begin{bmatrix} 8.1594 & 10.7420 \\ 10.7420 & 15.6500 \end{bmatrix}$$

$d \times 1$

$$\boldsymbol{\theta} = \begin{bmatrix} a \\ b \end{bmatrix}$$

$$\hat{\boldsymbol{\theta}}_N = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \|Y - \Phi\boldsymbol{\theta}\|_2^2$$

$$= (\Phi^\top\Phi)^{-1}\Phi^\top Y = \begin{bmatrix} -0.6839 \\ 0.2920 \end{bmatrix}$$

$$\hat{y}(11|10) = -ay(10) + bu(10)$$

$$= 0.6839 * 1.5 + 0.2920 * 2 = 1.61$$

Outline

1. Introduction to identification
2. Experiment design and model selection
3. Identification criterion
4. **Model validation**
5. A priori identifiabilty

Model Selection

We saw so far that output error identification is based on a fixed-model class:

$$\mathcal{M}(\boldsymbol{\theta}) = \{M(\boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^d\} \qquad \widehat{\boldsymbol{\theta}}_N = \arg\min_{\boldsymbol{\theta}} J_N(\boldsymbol{\theta})$$

$$J_N(\boldsymbol{\theta}) = \frac{1}{N}\sum_{t=1}^{N}\left(y(t) - \hat{y}(t|t-1;\boldsymbol{\theta})\right)^2$$

$\boldsymbol{\theta} \in \mathbb{R}^d$

This number is **fixed** and based on the model structure we have chosen.

To verify if our choice is correct, we need to test the performance of the model. But how?

Naive approach

Solve a sequence of identification problems for different model structures m and take the one that gives a minimum cost $J_I\left(\widehat{\boldsymbol{\theta}}_N^m\right)$

Procedure:

For $m = 1$ to m_{end}

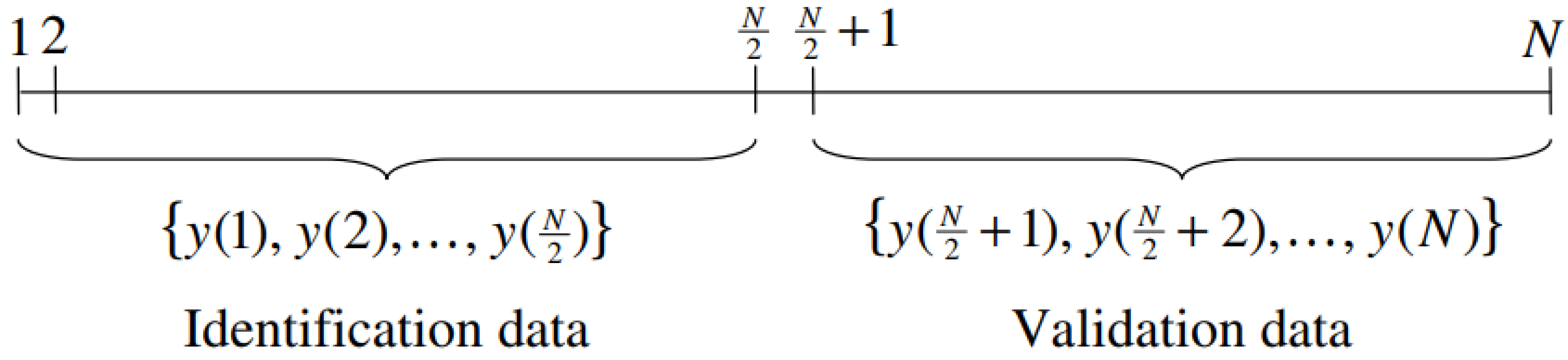
Find $\widehat{\boldsymbol{\theta}}_N^m$ by minimizing $J_I(\boldsymbol{\theta}^m) = \frac{1}{N}\sum_{t=1}^{N}\left(y(t) - \hat{y}(t;\boldsymbol{\theta}^m)\right)^2$

Choose the model m which gives the minimum value of $J_I\left(\widehat{\boldsymbol{\theta}}_N^m\right)$

This approach does not work! The value of $J_I\left(\widehat{\boldsymbol{\theta}}_N^m\right)$ will be always the minimum for any model m in that data set, but this does not imply that a different/more complex model better describes a system.

Approach 2: Validation

Suppose we have N data. We divide data in 2 subsequences:



Denote the cost on **identification data** as $J_I(\theta)$, and the cost on **validation data** as $J_V(\boldsymbol{\theta})$.

We want to choose the model that performs best on $J_V(\boldsymbol{\theta})$

Validation

The idea of a **validation** set is to **estimate the model performance** out of sample

1. Remove a **subset** from the training data

2. Identify the model on the remaining training data → the model will be trained on *less* data

3. Evaluate the model performance on the **held-out** set → this is an unbiased estimation of the out-of-sample error

4. Retrain the model on **all** the data

Validation

Procedure:

For $m = 1$ to m_{end}

Find $\hat{\boldsymbol{\theta}}_N^m$ by minimizing $J_I(\boldsymbol{\theta}^m) = \frac{1}{N}\sum_{t=1}^{N}\left(y(t) - \hat{y}(t;\boldsymbol{\theta}^m)\right)^2$

Evaluate $\boldsymbol{J_V\left(\hat{\theta}_N^m\right) = \frac{1}{N}\sum_{t=1}^{N}\left(y(t) - \hat{y}\left(t;\hat{\theta}_N^m\right)\right)^2}$

Choose the model m which gives the minimum value of $\boldsymbol{J_V\left(\hat{\theta}_N^m\right)}$

- This way we evaluate the model performance on a **fresh set of data**.
- $\boldsymbol{J_V\left(\hat{\theta}_N^m\right)}$ misures the **real performance** of the model

Rule of thumb

Given the dataset $\mathcal{D} = \{(u(1), y(1)), \dots, (u(N), y(N))\}$

K points: **validation** → $\mathcal{D}_{val}$

$N - K$ points: **training** → $\mathcal{D}_{train}$

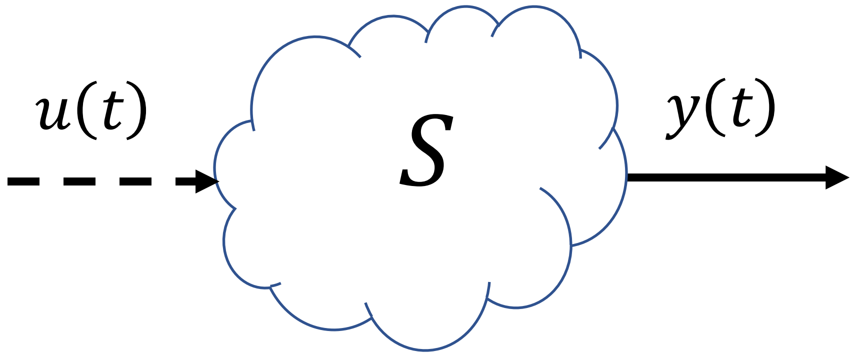
- **Small** K: bad estimate of $J_V(\boldsymbol{\theta})$
- **Large** K: possibility of identifying a bad model

Rule of thumb

$$K = \frac{N}{5}$$

The validation idea

The aim of identification is to estimate (learn) a model, which is a function that maps **inputs** u to **outputs** y



The identification process is performed by **minimizing a cost function** such as

$$J(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N}\left(y(t) - \hat{y}(t;\boldsymbol{\theta})\right)^2$$

on a certain dataset.

The ultimate goal is to learn a model that **performs well on unseen data**

- good approximation **out of sample**

The validation idea

- **<u>In-sample error:</u>** this is the error (as measured by the cost function or another index) that the model makes on **train data** (i.e. data used to estimate the parameters $\boldsymbol{\theta}$)

$$E_{in}(\boldsymbol{\theta}) = J(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N}\big(y(t) - \hat{y}(t;\boldsymbol{\theta})\big)^2$$

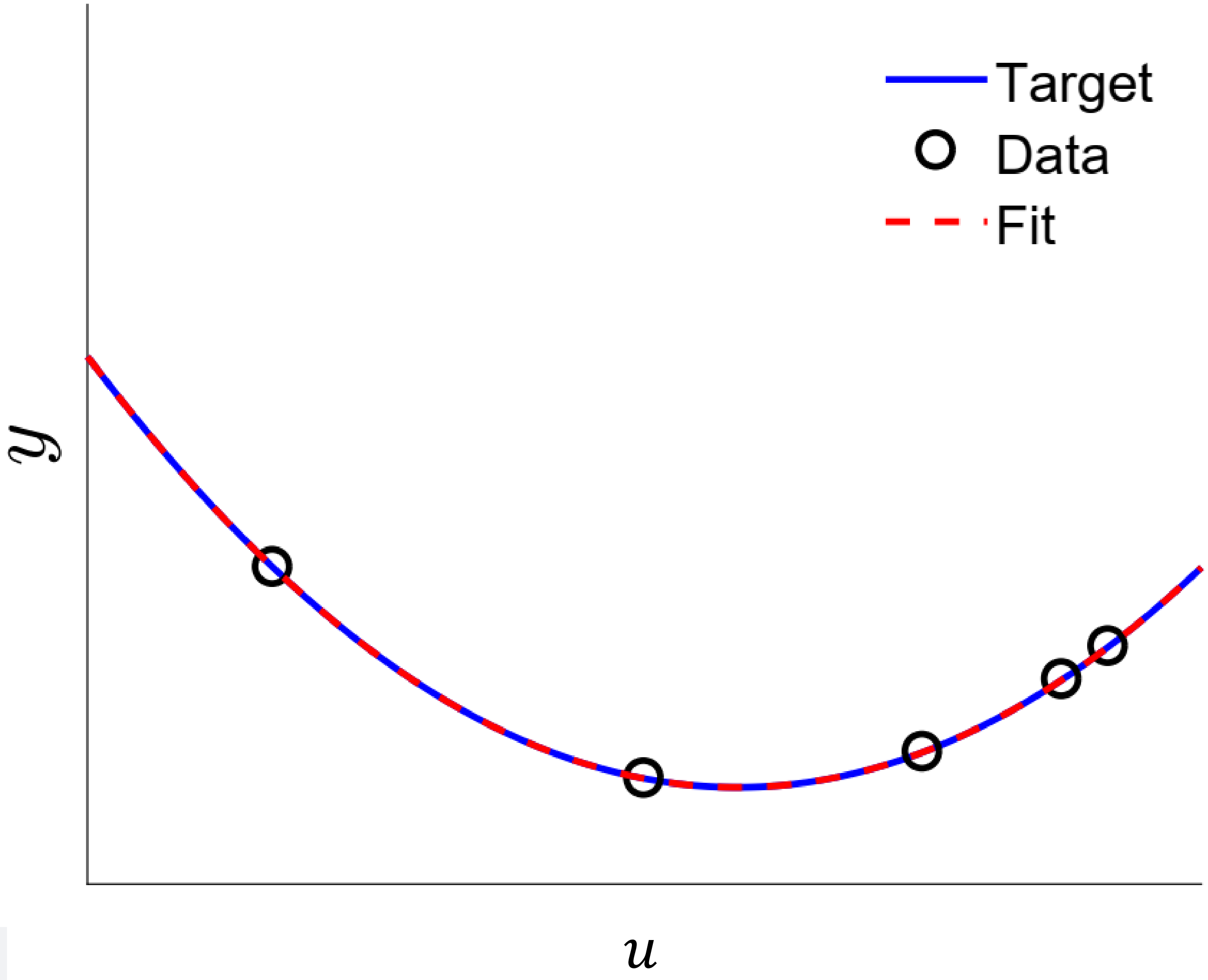
- **<u>Out-of-sample error:</u>** this is the error (as measured by the cost function or another index) that the model makes on **test data** $\mathcal{D}_{\boldsymbol{val}} = \{(\tilde{\boldsymbol{u}}(i), \tilde{y}(i))\}_{i=1}^{N_T}$ (i.e. data **NOT used** to estimate the parameters $\boldsymbol{\theta}$)

$$E_{out}(\boldsymbol{\theta}) = \frac{1}{N_T}\sum_{i=1}^{N_T}\big(\tilde{y}(i) - \hat{y}(t;\boldsymbol{\theta})\big)^2$$

Overfitting and noise on outputs

- Example: Simple system to identify
- $N = 5$ points
- Model: perfectly fit the data and the real system output

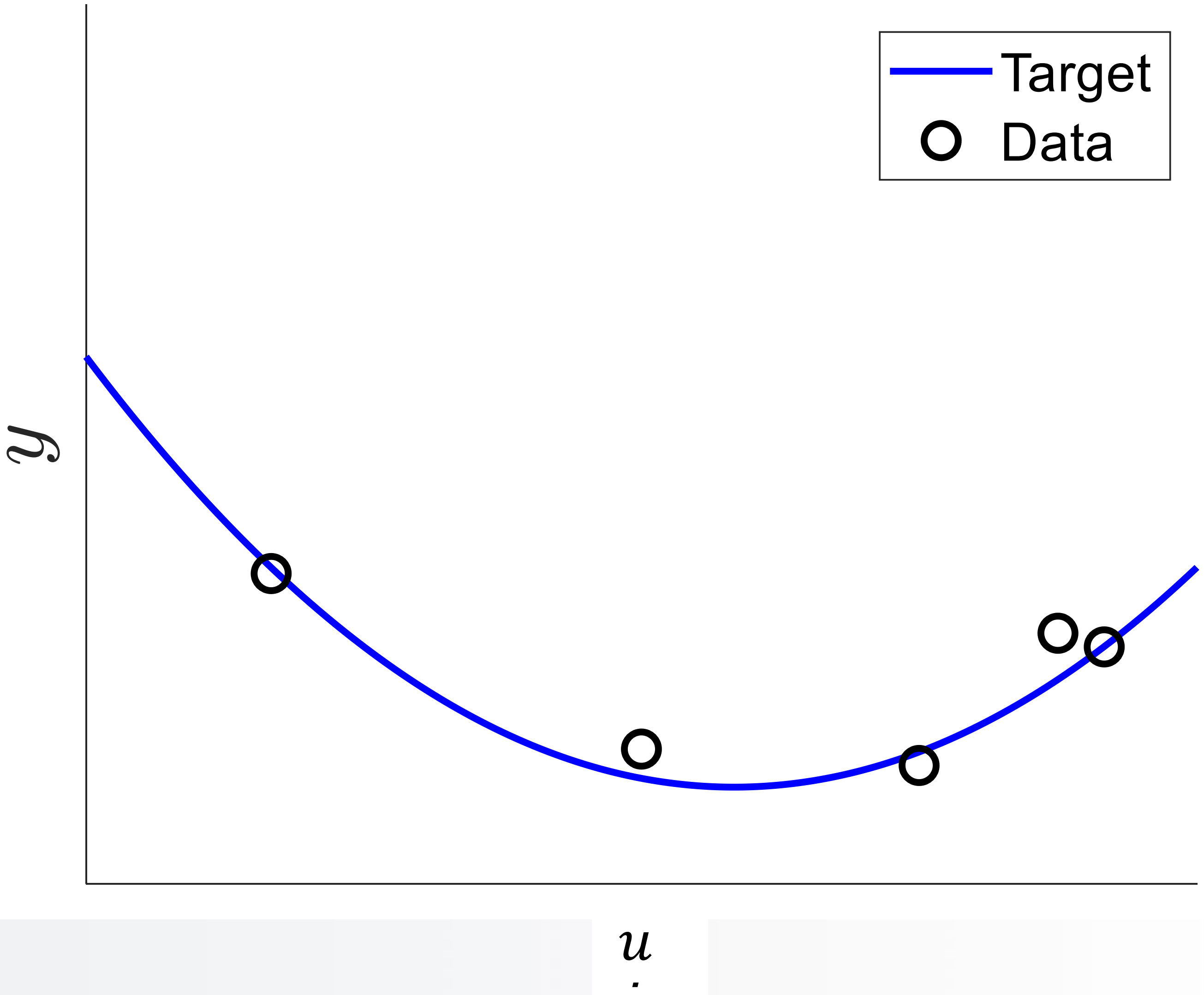
$E_{in} = 0$ $E_{out} = 0$



Overfitting and noise on outputs

- Example: Simple system to identify

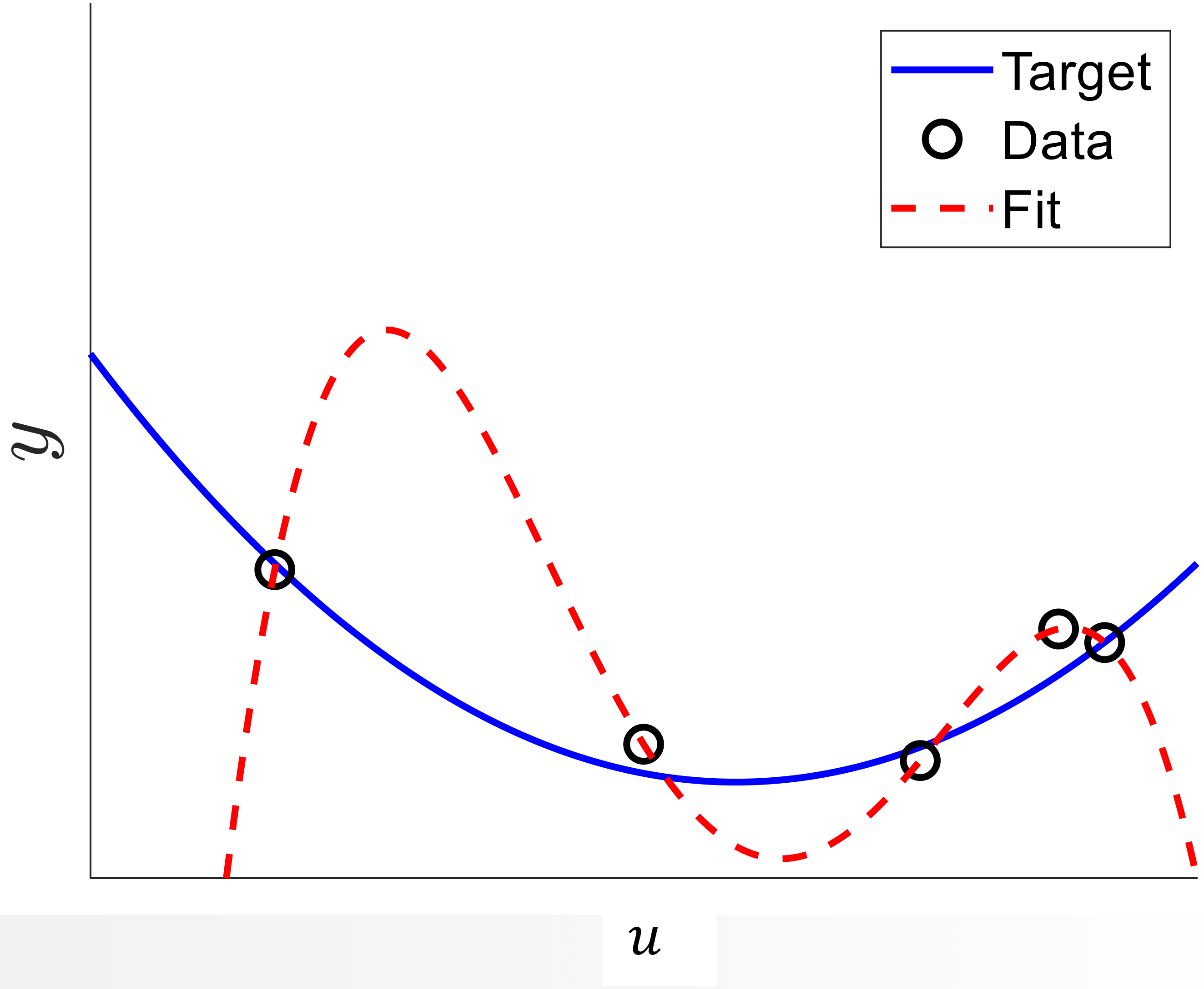
- $N = 5$ **noisy** points



Overfitting and noise on outputs

- Example: Simple system to identify
- $N = 5$ **noisy** points
- Model: perfectly fit the data, but not the real system

$$E_{in} = 0 \qquad E_{out} = \textbf{huge}$$



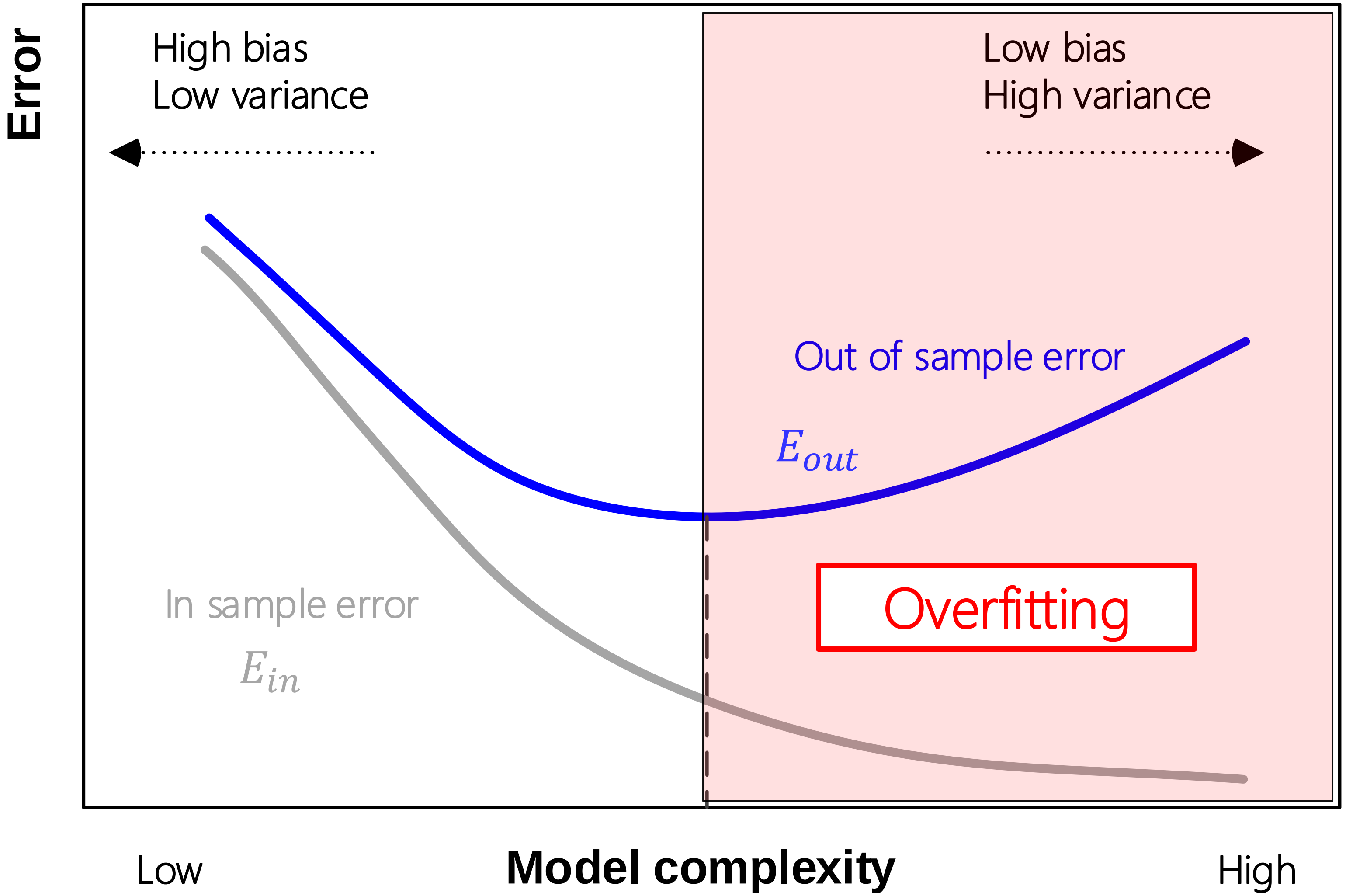
Overfitting vs. model complexity

We talk of **overfitting** when decreasing E_{in} leads to increasing E_{out}

Major source of failure for identification problems

Overfitting leads to bad generalization

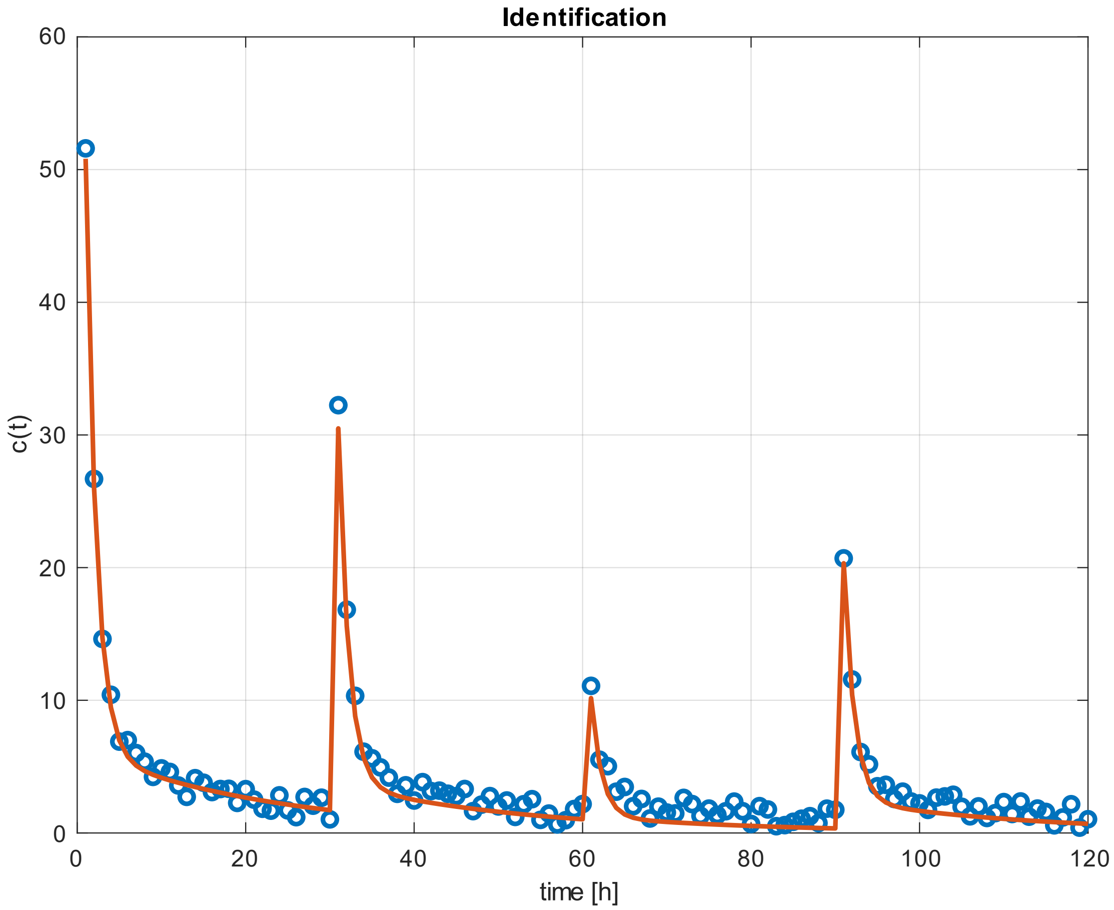
A model can exhibit bad generalization even if it does not overfit

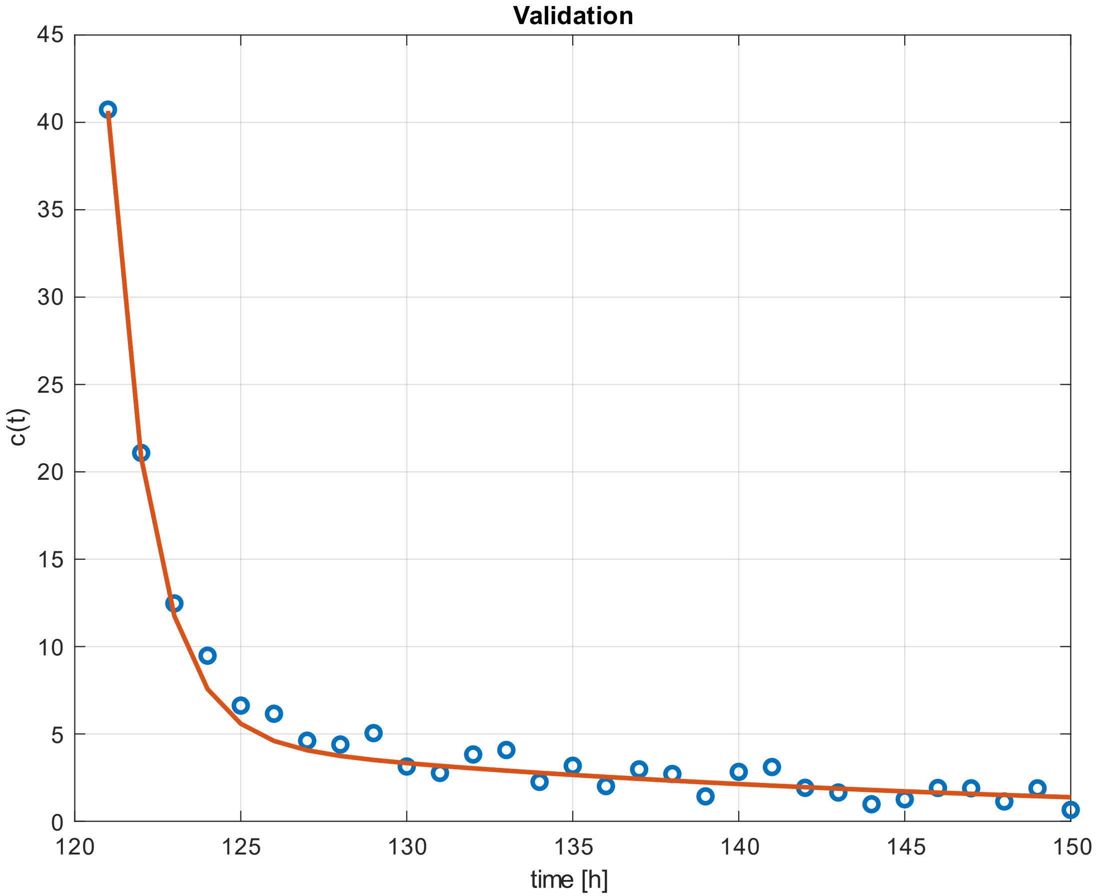


Worked example

$E_{in} = 0.35$

$E_{out} = 0.62$





$$\boldsymbol{\theta} = \{k_{01},\ k_{12},\ k_{21}, V_1\} = \{0.23; 0.26; 0.10; 9.84\}$$

Outline

1. Introduction to identification
2. Experiment design and model selection
3. Identification criterion
4. Model validation
5. **A priori identifiabilty**

A priori identifiability

Only parameters that meet certain conditions can be determined from input/output data. The parameter set can sometimes be determined uniquely, sometimes not.

Identifiability Problem:

1. determine whether it is possible to find 1 or more sets of solutions for the unknown parameters of the model, from data collected in experiments carried out on the real system.

2. Finding validity ranges for parameters of non-identifiable models

A priori identifiability

Identifiability analysis is a preliminary step in model analysis for parametric estimation.

From this analysis, the minimum conditions necessary to obtain unique estimates from the noisy and limited real data are obtained.

Goal: to establish theoretically whether, given the structure of the model and a certain configuration of inputs and outputs, it is possible to derive the unknown parameters of the model in the purely ideal case in which the model is error-free and the continuous-time outputs are known exactly.

A priori identifiability

Only if the model is identifiable a priori it does make sense to try to numerically estimate the value of its parameters from the experimental data

Remedies for a priori non-identifiability:

1) enrich the experiment, e.g. by adding measures;

2) reduce the complexity of the model, e.g. by reducing the number of compartments or parameters or adding constraints on the parameters.

Importance of a priori identifiability in the qualitative design of the experiment: determine minimum number of inputs and outputs that guarantee identifiability

A priori identifiability

It does NOT depend on the data a posteriori, but only on the a priori structure of the model.

The random nature of the actual data does NOT affect these results

There can be 3 possible cases

- Non-identifiability
- Generic identifiability
- Univocal identifiability

Non-identifiability VS Identifiability

A p parameter is said to be **NON-IDENTIFIABLE** in the interval $[t_0, T]$ if there is an **INFINITE** number of solutions.

- If a model has even one **NON-IDENTIFIABLE** parameter, then the entire structure is said to be **NON-IDENTIFIABLE**.

A p parameter is said to be **IDENTIFIABLE** in the interval $[t_0, T]$ if there exists a **FINITE** number of solutions (different from the null one).

- If all parameters are **IDENTIFIABLE** parameters, then the entire structure is said to be **IDENTIFIABLE**.

Univocal Identifiability

A p parameter is said to be **UNIVOCALLY IDENTIFIABLE** in the interval $[t_0, T]$ if there exists a **ONE AND ONLY ONE** solution (different from the null one).

- If all parameters are **UNIVOCALLY IDENTIFIABLE** parameters, then the entire structure is said to be **UNIVOCALLY IDENTIFIABLE**.

- If even just one parameter is **NOT UNIVOCALLY IDENTIFIABLE** , then the entire structure is said to be **NOT UNIVOCALLY IDENTIFIABLE**.

Necessary condition for Identifiability

Necessary condition for a priory identifiability is that the system is input-output connectable.

- A compartment is input connectable if it can be reached by any input compartment.
- A compartment is output connectable if any output compartment can be reached from it.
- A system is **input-output connectable** if every compartment of the system is input-output connectable (any compartment is reachable from one input and it is connected to at least one output).

Remark

A trick to verify a priori identifiability is to study the transfer function.

Discrete time:

$$G(z,\boldsymbol{\theta}) = C(\boldsymbol{\theta})(zI - A(\boldsymbol{\theta}))^{-1}B(\boldsymbol{\theta}) + D(\boldsymbol{\theta})$$

$$y(t) = G(z,\boldsymbol{\theta})u(t) = \frac{b_0 + b_1 z^{-1} + \cdots + b_p z^{-p}}{1 - a_1 z^{-1} - \cdots - a_m z^{-m}} u(t)$$

In both cases the parameters of the transfer function can be expressed as a function of the parameters of the state space model.

Continuous time:

$$G(s,\boldsymbol{\theta}) = C(\boldsymbol{\theta})(sI - A(\boldsymbol{\theta}))^{-1}B(\boldsymbol{\theta}) + D(\boldsymbol{\theta})$$

$$y(t) = G(s,\boldsymbol{\theta})u(t) = \frac{b_0 s^p + b_1 s^{p-1} + \cdots + b_{p-1} s + b_p}{s^m + a_1 s^{m-1} + \cdots + a_{m-1} s + a_m} u(t)$$

Remark

Let's consider the pharmacokinetic models case:

Continuous time:

$$y(t) = G(s,\boldsymbol{\theta})u(t) = \frac{b_0 s^p + b_1 s^{p-1} + \cdots + b_{p-1}s + b_p}{s^m + a_1 s^{m-1} + \cdots + a_{m-1}s + a_m} u(t)$$

The coefficients (a_l, b_l) are function of the parameters of the state space model (for instance (k_{ij}, V_i)).

Since $u(t)$ is known and $y(t)$ is the system response, then (a_l, b_l) are identifiable.

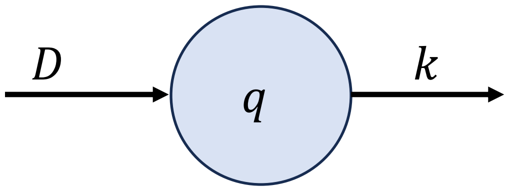
We can write down the system of algebraic relations between (a_l, b_l) and (k_{ij}, V_i). This system is called **exhaustive summary.**

If it admits a solution, then the system is **a priori identifiable.**

$$G(s,\boldsymbol{\theta}) = C(\boldsymbol{\theta})(sI - A(\boldsymbol{\theta}))^{-1}B(\boldsymbol{\theta}) + D(\boldsymbol{\theta})$$

Example 1

Consider the system



$$\dot{q}(t) = -kq(t) + D\delta(t) \qquad q(0) = 0 \qquad \boldsymbol{\theta} = \{k, V\}$$

$$q(t) = De^{-kt}$$

$$c(t) = \frac{q(t)}{V}$$

Output of the model

Whose transfer function is

$$G(s, \boldsymbol{\theta})u(t) = \frac{\frac{1}{V}}{s+k} = \frac{b}{s+a}$$

The exhaustive summary is

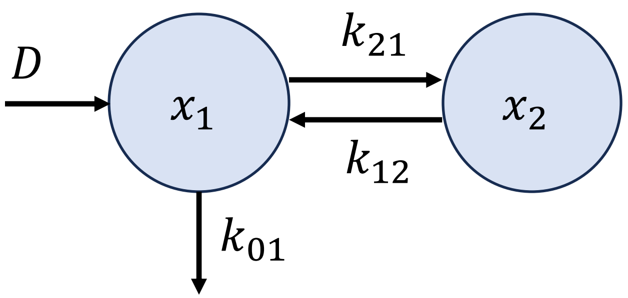
$$b = \frac{1}{V} \Rightarrow V = \frac{1}{b}$$

$$a = k$$

The system is univocally identifiable.

Example 2

Consider the system



$$\dot{x}_1(t) = -(k_{01} + k_{21})x_1(t) + k_{12}x_2(t) + D\delta(t)$$

$$\dot{x}_2(t) = k_{21}x_1(t) - k_{12}x_2(t)$$

$$y(t) = \frac{x_1(t)}{V_1}$$

$$\boldsymbol{\theta} = \{k_{01},\ k_{12},\ k_{21}, V_1\}$$

Whose transfer function is

$$G(s,\boldsymbol{\theta})u(t) = \frac{\frac{s + k_{12}}{V_1}}{s^2 + (k_{01} + k_{12} + k_{21})s + k_{01}\ k_{12}} = \frac{b_1 s + b_0}{s^2 + a_1 s + a_2}$$

The exhaustive summary is

$$b_1 = \frac{1}{V_1}$$
$$b_0 = \frac{k_{12}}{V_1}$$
$$a_2 = k_{01}\ k_{12}$$
$$a_1 = k_{01} + k_{12} + k_{21}$$

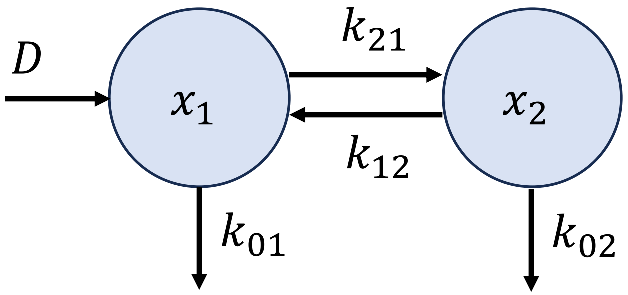
$\Rightarrow$

$$V_1 = \frac{1}{b_1}$$
$$k_{12} = b_0 V_1$$
$$k_{01} = \frac{k_{12}}{a_2}$$
$$k_{21} = a_1 - k_{01} - k_{12}$$

The system is univocally identifiable.

Example 2

Consider the system



$$\dot{x}_1(t) = -(k_{01} + k_{21})x_1(t) + k_{12}x_2(t) + D\delta(t)$$
$$\dot{x}_2(t) = k_{21}x_1(t) - (k_{02} + k_{12})x_2(t)$$
$$y(t) = \frac{x_1(t)}{V_1}$$

$$\boldsymbol{\theta} = \{k_{01}, k_{02}, k_{12},\ k_{21}, V_1\}$$

Whose transfer function is

$$G(s,\boldsymbol{\theta})u(t) = \frac{\frac{s + k_{12} + k_{02}}{V_1}}{s^2 + (k_{01} + k_{02} + k_{12}+k_{21})s + (k_{01}\ k_{12} + k_{01}\ k_{02} + k_{21}\ k_{02})} = \frac{b_1 s + b_0}{s^2 + a_1 s + a_2}$$

The exhaustive summary is

$$b_1 = \frac{1}{V_1} \quad \Rightarrow \quad V_1 = \frac{1}{b_1}$$

$$b_0 = \frac{k_{12} + k_{02}}{V_1}$$

$$a_2 = k_{01}\ k_{12} + k_{01}\ k_{02} + k_{21}\ k_{02}$$
$$a_1 = k_{01} + k_{02} + k_{12}+k_{21}$$

4 equations and 5 unknown variables.

The system is non identifiable.

UNIVERSITÀ
DEGLI STUDI
DI BERGAMO

Dipartimento
di Ingegneria Gestionale,
dell'Informazione e della Produzione