Aalto University
School of Electrical Engineering

ELEC-E8103 Modelling, Estimation and Dynamic Systems

System Identification

Houari Bettahar

Department of Electrical Engineering and Automation

Aalto University, School of Electrical Engineering

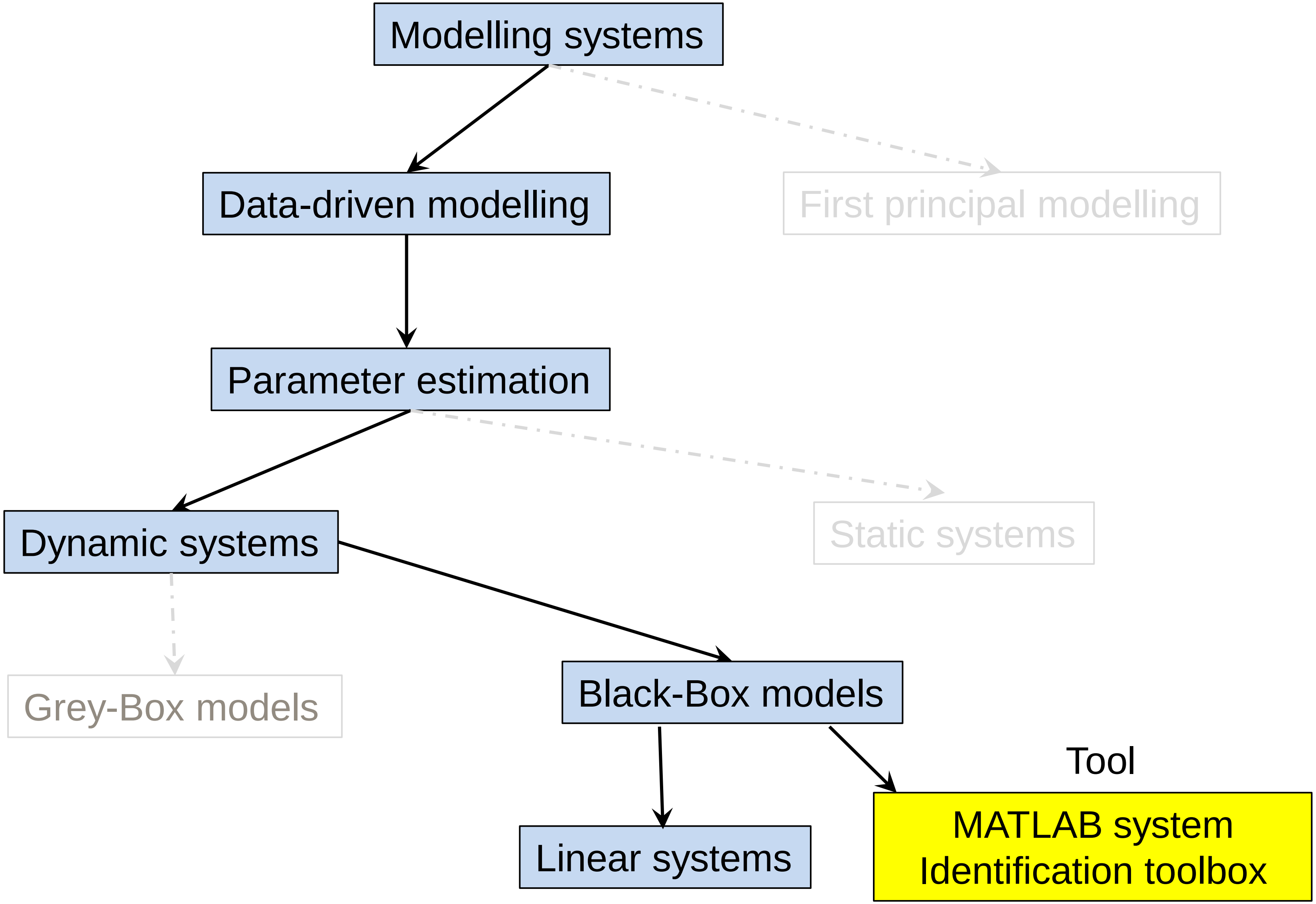
Email: firstname.lastname@aalto.fi

Learning goals

Course Learning Outcomes

- Select proper modeling approach for specific practical problems,
- Formulate mathematical models of physical systems,
- Construct models of systems using modeling tools such as MATLAB and Simulink,
- Estimate the parameters of linear and nonlinear static systems from measurement data,
- Identify the models of linear dynamic systems from measurement data

Recap


Modelling systems
Data-driven modelling
First principal modelling
Parameter estimation
Dynamic systems
Static systems
Grey-Box models
Black-Box models
Tool
Linear systems
MATLAB system
Identification toolbox

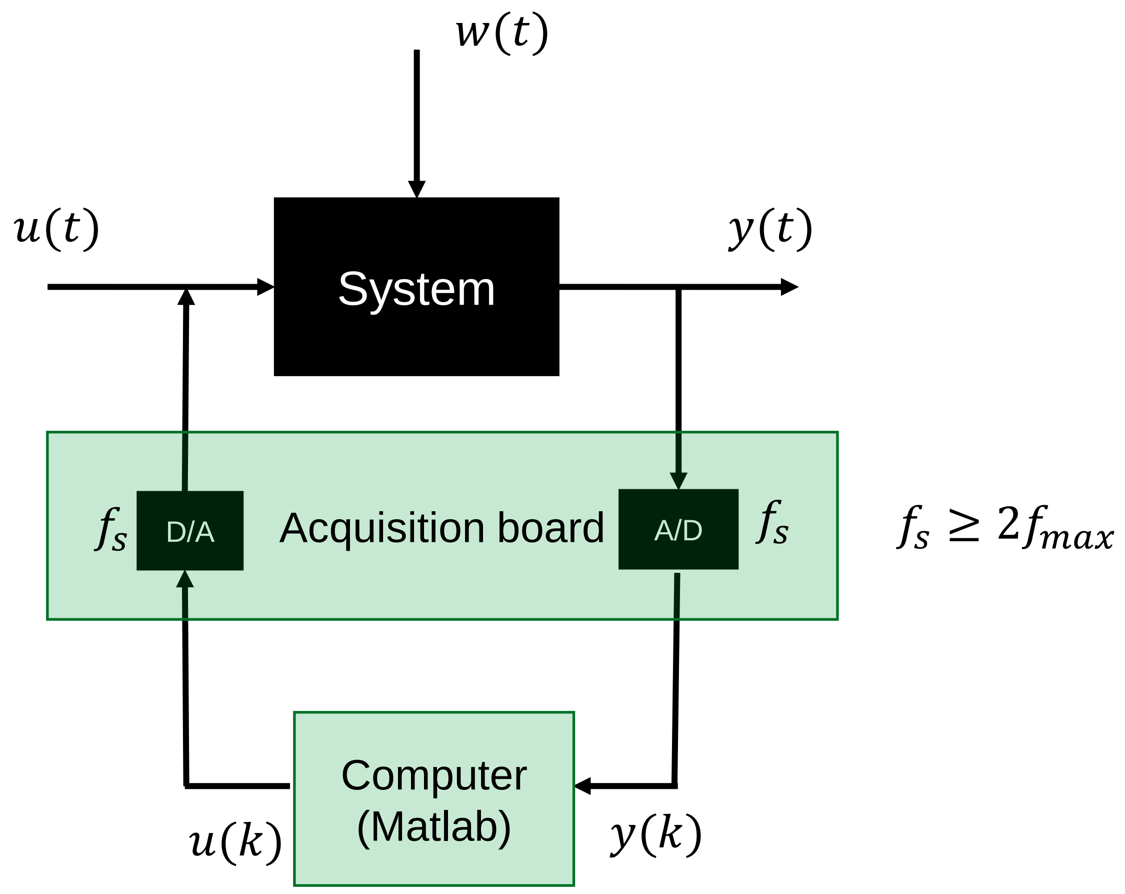
Learning goals

- Learn the procedure of System Identification through Matlab Toolbox for system modeling
  - Experimental design and data collection for modelling
  - Model structure and order
  - Parameter estimation
  - Model validation

Prior knowledge

- Signal processing: filtering, detrend….
- Cross-correlation, autocorrelation
- Linear dynamic system: Bode diagram, zeros and poles

Experimental design and data collection for modelling



Experimental design

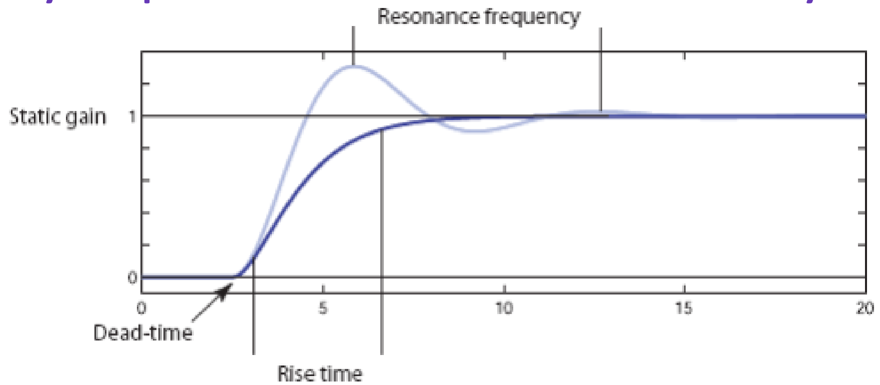
Experimental design and data collection for modelling

- A two-stage approach.

1. Preliminary experiments:
2. Data collection for model estimation:

Experimental design and data collection for modelling

- 1. Preliminary experiments: Transient analysis



Useful for obtaining qualitative information about system:

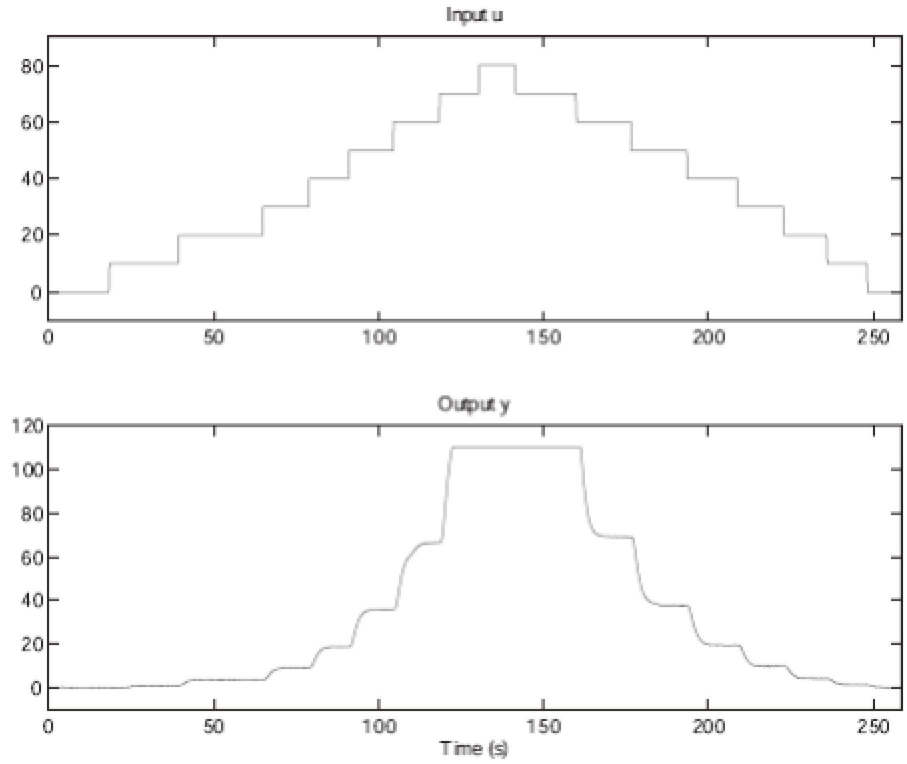
- indicates dead-times (delay), static gain, rise time
- aids sampling time selection (rule-of-thumb: 4-10 samples per rise time)
  - Sampling that is considerably faster than the system dynamics leads to data redundancy
  - Sampling that is considerably slower than the system dynamics leads to serious difficulties in determining the parameters that describe the dynamics

Aalto University
School of Electrical Engineering

Experimental design and data collection for modelling

1. Preliminary experiments:

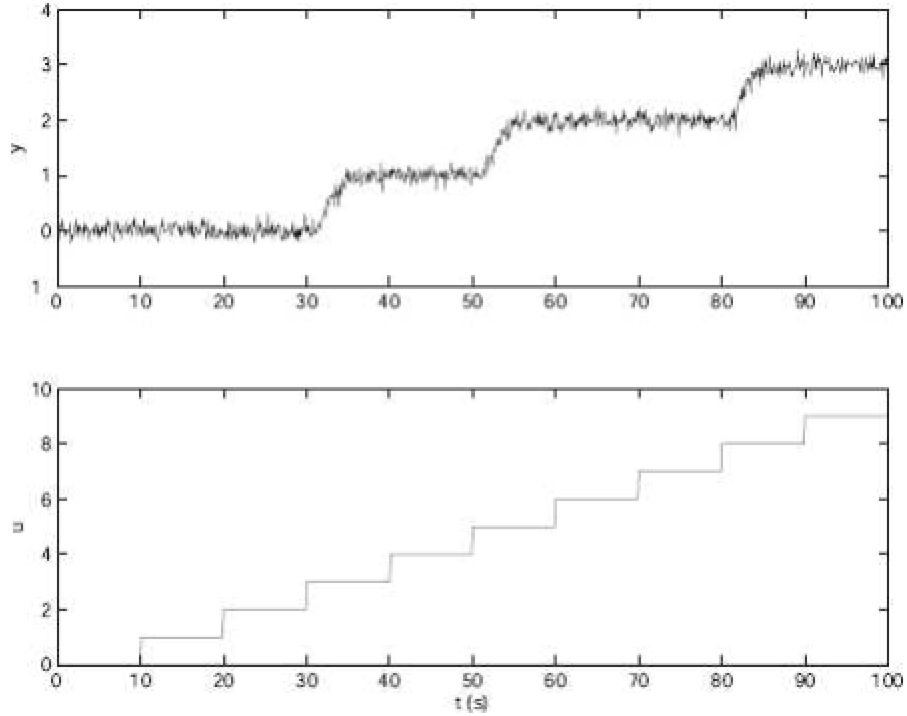
- test linearity by a sequence of step response tests



Experimental design and data collection for modelling

1. Preliminary experiments:

- Friction can be detected by using small step increases in input

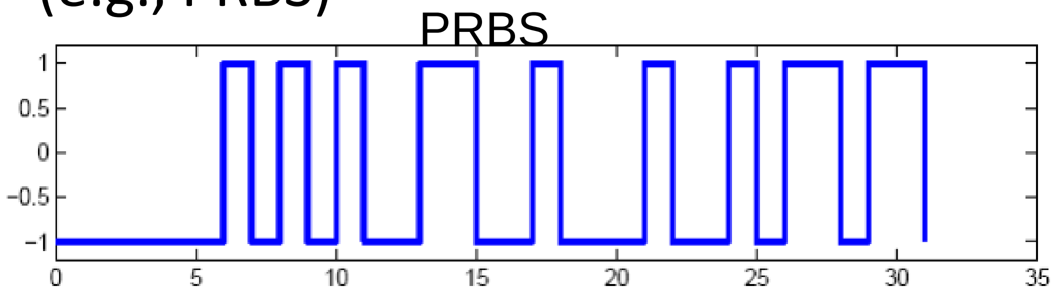


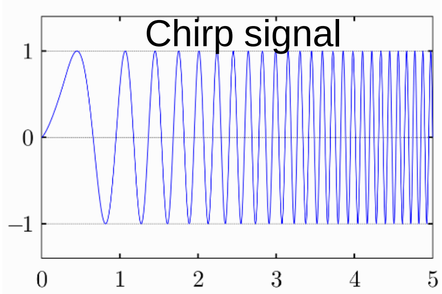
Output moves every two or three steps.

Experimental design and data collection for modelling

2. Data collection for model estimation:

- Input signal should excite all relevant frequencies
- good choice is often a binary sequence with random hold times (e.g., PRBS)
-



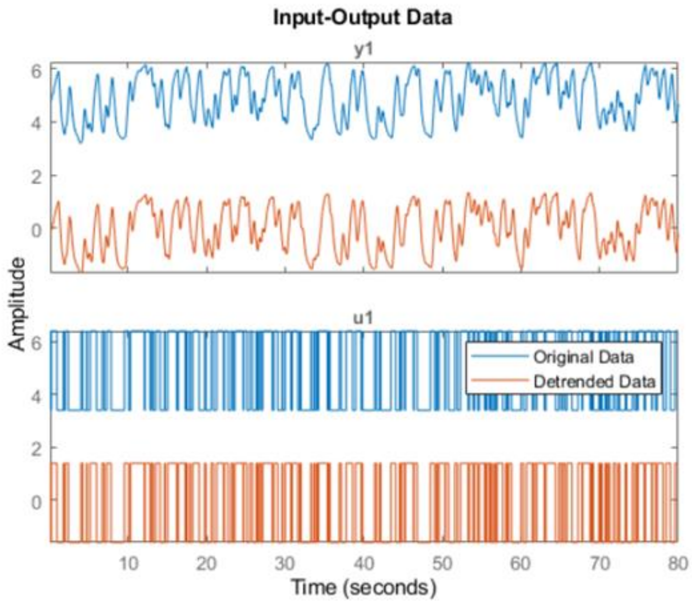


- Chirp signal → works very well for nonlinear systems
- Trade-off in selection of signal amplitude
  - large amplitude gives high signal-to-noise ratio, low parameter variance

Experimental design and data collection for modelling

Preprocessing of data

- Preproscessing by removing undesired information
  - Mean of the input and output data



Aalto University
School of Electrical
Engineering

Experimental design and data collection for modelling

Preprocessing of data

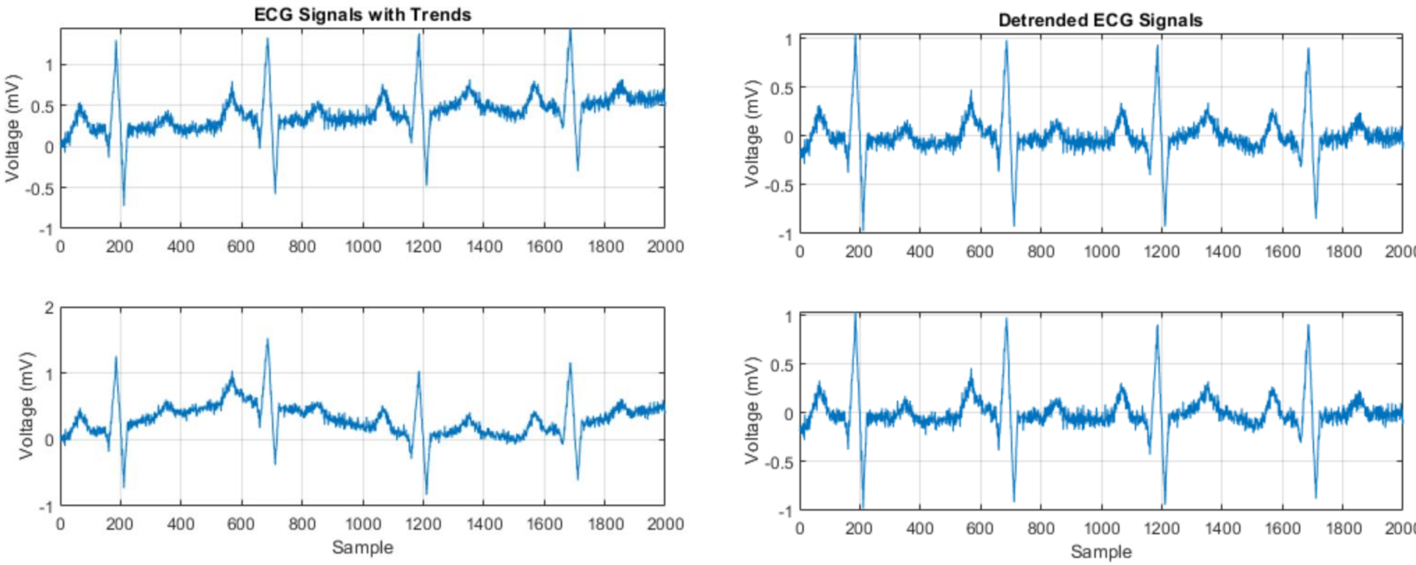
Trend

Trend

Detrend

Linear

nonlinear



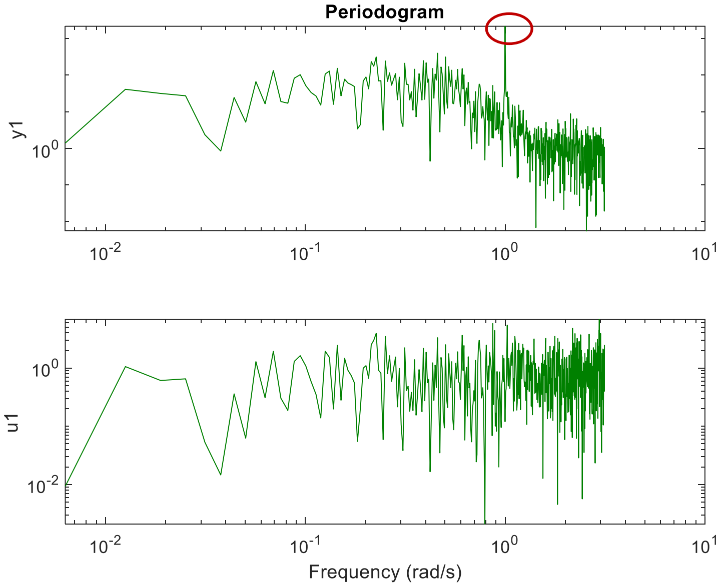
Outlier

- Obvious error data, most obvious in residual
- Remove by hand or algorithm

Experimental design and data collection for modelling

Preprocessing of data

- Removing disturbances
  - Low frequency distrubances
    - High pass filter
  - High frequency distrubances
    - Low-pass filter
  - Distrubance at certain frequency
    - Stop-band filter



Learning goals

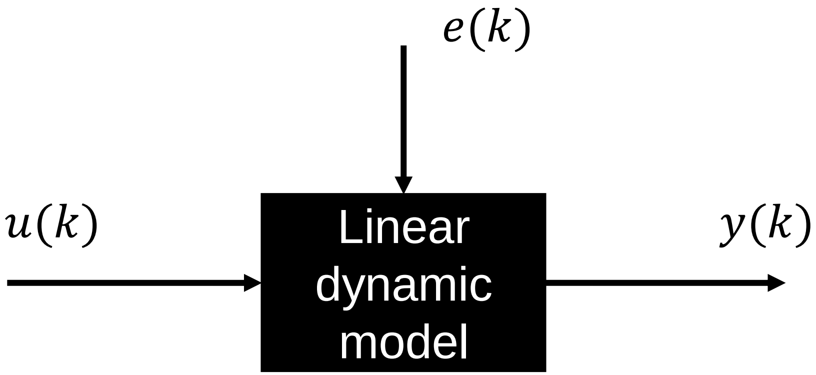
- Learn the procedure of System Identification through Matlab Toolbox for system modeling
  - Experimental design and data collection for modelling
  - ➢ Preliminary experiments
  - ➢ Data collection and data preprocessing
  - Model structure and order: ARX, ARMAX, OE, BJ
  - Parameter estimation
  - Model validation

Model structure and order

- **Do we know anything about the system apriori?**
  - **Black-Box model:** Flexible structure, is a method for the development of **models** based on process data.
  - **Grey-Box models:** Tailor-made structures, made to incorporate prior knowledge: Structured differential equation with some parameters unknown
- **Is the output a linear or a nonlinear function of the input?**
- **Do we want to describe also how disturbances affect the output ?**

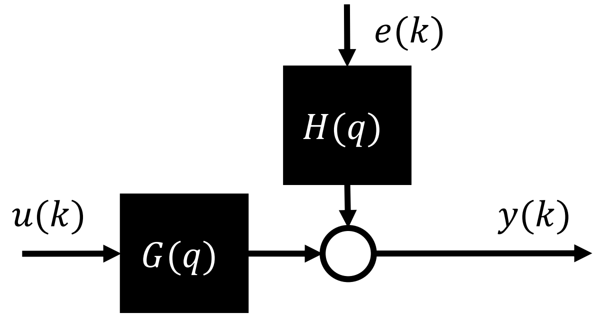
Model structure and order

- Linear dynamic model, with components describing the relation of
  - $y(k)$ and $u(k)$
  - $y(k)$ and $e(k)$
- Before estimate the parameters, we need to select
  - The model structure
  - The model order



Model structure and order

Polynomial models



- A linear time discrete model can be written as

$$y(k) = G(q,\theta)u(k) + H(q,\theta)e(k)$$

  - $e(k)$ is the white noise, assuming $w(k) = H(q,\theta)e(k)$
  - θ is the model parameter
  - q is the shift operator, $y(k) = qy(k-1), y(k-1) = q^{-1}y(k)$

$$G(q,\theta) = \frac{q^{-n_k}(b_1 + b_2 q^{-1} + \cdots + b_{n_b} q^{-n_b+1})}{1 + f_1 q^{-1} + \cdots + f_{n_f} q^{-n_f}} = \frac{B(q)}{F(q)}$$

$$H(q,\theta) = \frac{1 + c_1 q^{-1} + \cdots + c_{n_c} q^{-n_c}}{1 + d_1 q^{-1} + \cdots + d_{n_d} q^{-n_d}} = \frac{C(q)}{D(q)}$$

- The parameter vector θ contains the coefficient $\{b_k\}, \{f_k\}, \{c_k\}, \{d_k\}$
- The different variations of the model is the so-called model structure
- n_b, n_c, n_d, n_f, determines the order of the polynomial transfer function
- n_k determines the time delay

Model structure and order

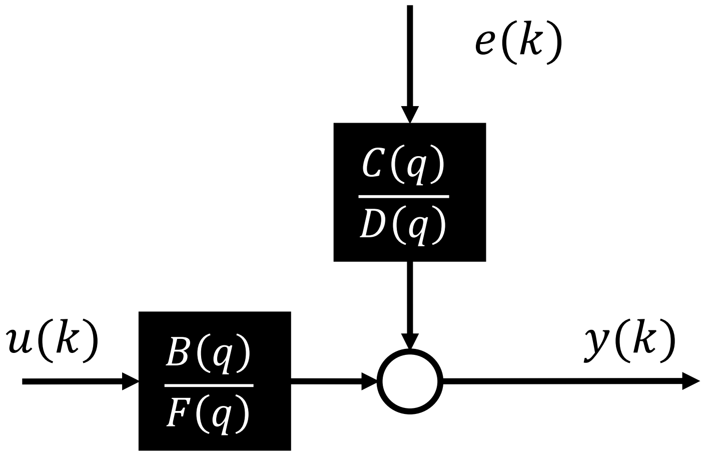
Polynomial models

- Flexibility: Approximates a wide range of functions, simple to complex, including linear, quadratic, cubic, and higher-order relationships.
- Nonlinearity: Captures nonlinear behavior in real-world systems.
- Simplicity: Simple and computationally efficient, ideal for practical applications with limited computational resources.
- Interpretability: Easy to interpret, especially with low-degree polynomials, aiding understanding of system relationships.

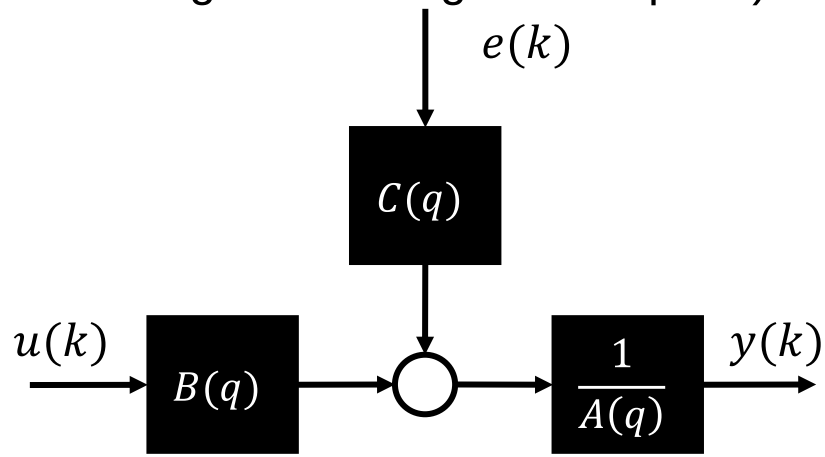
Model structure and order

Common model structures

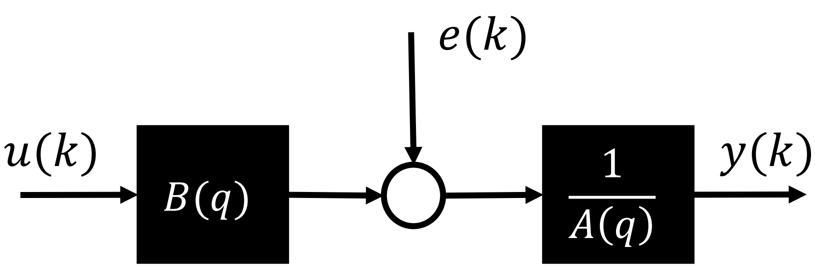
BJ (Box-Jenkins)



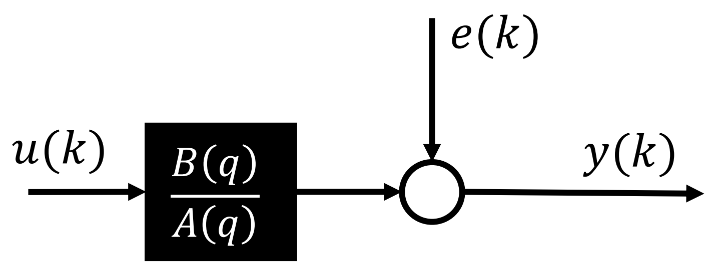
ARMAX (AutoRegressive-Moving-Average with eXogenous inputs)



ARX (AutoRegressive with eXogenous inputs)



OE (Output Error)

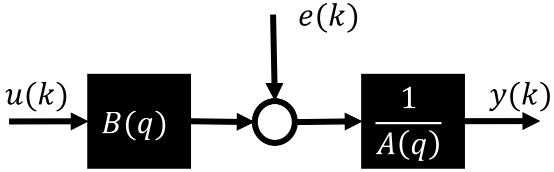


Model structure and order

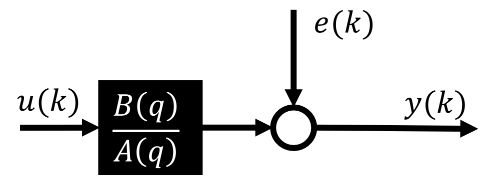
Why do we need to have a strategy for model structure selection?

Let's assume we will have maximum 3rd order polynomial

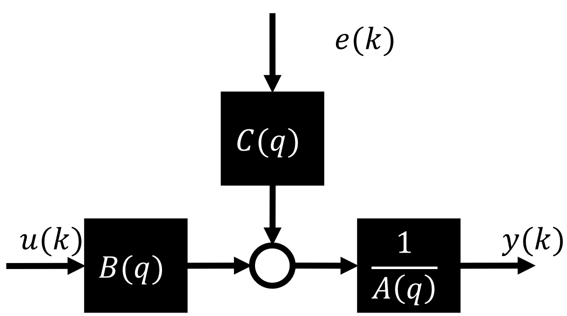
ARX: 9 options



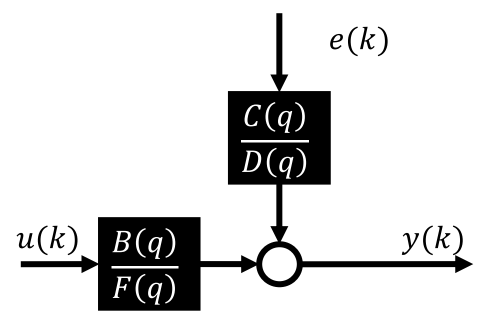
OE: 9 options



ARMAX: 27 options



BJ: 81 options



Model structure and order

Drawbacks of higher-order system selection

- Overfitting:
  - Captures noise, not underlying system dynamics.
  - Performs poorly on unseen data due to memorized noise.
- Increased Computational Complexity:
  - Needs more data for accurate parameter estimation.
  - Challenging in data collection, especially for complex models (expensive)
- Sensitivity to Noise:
  - Small input fluctuations cause significant prediction variations.
- Limited Robustness:
  - Sensitive to system changes and disturbances.
  - Less robust compared to simpler models.

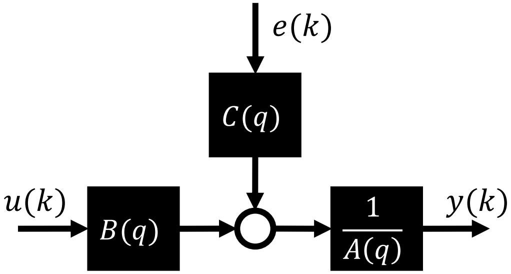
Model structure and order

Drawbacks of inappropriate model structure selection:

- Leads to inaccurate representations of system behavior.
- Results in poor predictions and reduced model effectiveness.
- Obstruct understanding of the underlying patterns in the data.
- Obstruct achieving desired performance outcomes.

Model structure and order

Comments on ARMAX



- Model for both input and noise, which enters the system early

$$A(q)y(k) = B(q)u(k) + C(q)e(k)$$

- Has interesting special cases
  - Autoregressive (AR) model: uses a linear combination of past observations to predict future values

$$A(q)y(k) = e(k)$$

$$\left(1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a}\right)y(k) = e(k)$$

  - Moving average (MA) model: calculates the average of a subset of the data points within a specified window of time. This window "moves" through the data set as new observations become available

$$y(k) = C(q)e(k) = \left(1 + c_1 q^{-1} + \cdots + c_{n_c} q^{-n_c}\right)e(k)$$

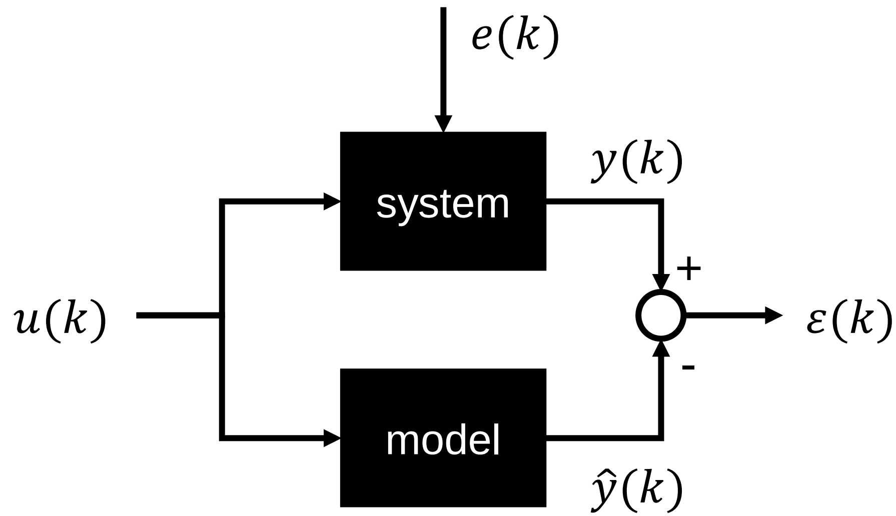
  - ARMA model: is a combination of the Autoregressive (AR) model and the Moving Average (MA) model.

$$A(q)y(k) = C(q)e(k)$$

Parameter estimation

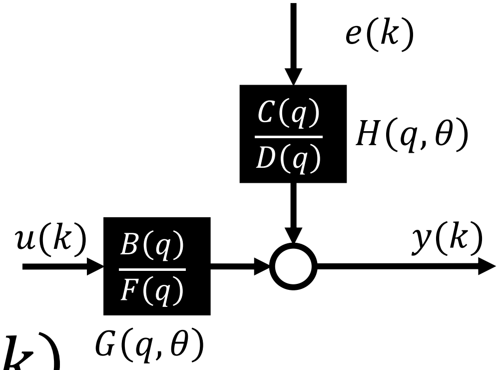
Parameter estimation of model structures based on prediction error

- Goal
  - For a given model structure
  - Find the best parameter set θ
    - $\varepsilon(k) = y(k) - \hat{y}(k)$ is somehow minimized
    - $\hat{y}(k)$ is the model prediction
- How to calculate the prediction?
  - Known conditions:
    - Input values
    - Previous output values



Parameter estimation

Prediction



- For linear time discrete model

$$y(k) = G(q,\theta)u(k) + H(q,\theta)e(k)$$

- Multiply both side by $H^{-1}(q,\theta)$ and rearrange, we have

$$y(k) = \left(1 - H^{-1}(q,\theta)\right)y(k) + H^{-1}(q,\theta)G(q,\theta)u(k) + e(k)$$

Notice, the noise term is white.

- The prediction of $y(k)$ is

$$\hat{y}(k|\theta) = \left(1 - H^{-1}(q,\theta)\right)y(k) + H^{-1}(q,\theta)G(q,\theta)u(k)$$

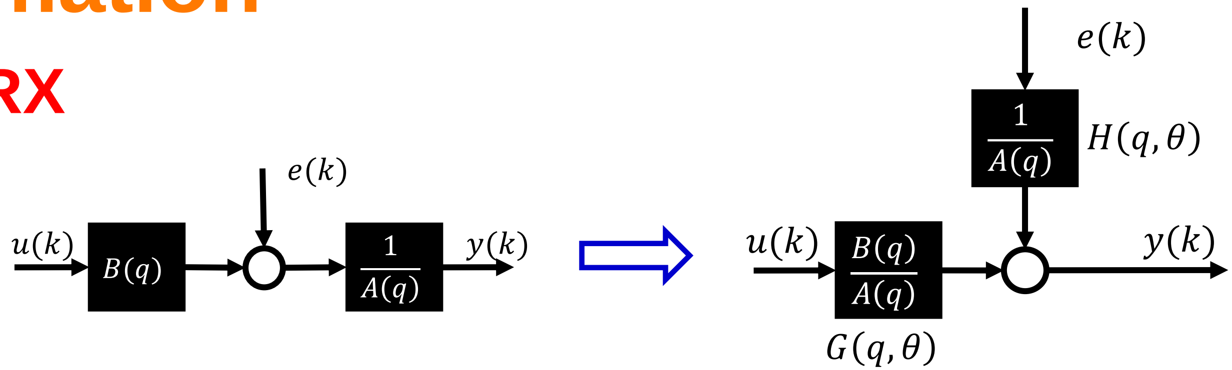
So we can use the $u(k)$ and old $y(k)$ to predict new $y(k)$

$\hat{y}(k|\theta) = \left(1 - H^{-1}(q,\theta)\right)y(k) + H^{-1}(q,\theta)G(q,\theta)u(t)$

Parameter estimation

Prediction using ARX

- The prediction is



$$\hat{y}(k) = \underbrace{-(a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a})}^{1 - H^{-1}(q,\theta)} y(k) + \overbrace{q^{-n_k}\left(b_1 + b_2 q^{-1} + \cdots + b_{n_b} q^{-n_b+1}\right)}^{H^{-1}(q,\theta)G(q,\theta)} u(k)$$

- In a more compact form: $\hat{y}(k|\theta) = \theta^T \varphi(k)$
- where

$$\theta = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{n_a} \\ b_1 \\ \vdots \\ b_{n_b} \end{bmatrix} \qquad \varphi(k) = \begin{bmatrix} -y(k-1) \\ -y(k-2) \\ \vdots \\ -y(k-n_a) \\ u(k-n_k) \\ \vdots \\ u(k-n_k-n_b+1) \end{bmatrix}$$

Parameter estimation

$$SSE(\beta) = (Y - X\beta)^T(Y - X\beta)$$
$$\hat{\beta} = (X^TX)^{-1}X^TY$$

Parameter estimation using Linear Regression

- If the problem is a linear regression, e.g. in the case of ARX

$$\hat{y}(k|\theta) = \theta^T\varphi(k) = \varphi(k)^T\theta$$

- The error for sample k is:

$$\varepsilon(k,\theta) = y(k) - \varphi(k)^T\theta$$

- The loss function is:

$$V_N(\theta) = \frac{1}{N}\sum_{k=1}^{N}\varepsilon^2(k,\theta) = \frac{1}{N}\varepsilon_N^T\varepsilon_N = \frac{1}{N}(y_N - \varphi_N\theta)^T(y_N - \varphi_N\theta)$$

  – $\varepsilon_N, y_N, \varphi_N$ are column vectors of size N of $\varepsilon(k,\theta)$, $y(k)$ and $\varphi(k)^T$.

- The best estimation $\hat{\theta}$ is found if $\frac{d}{d\theta}V_N(\theta) = 0$, and we have

$$\hat{\theta} = ({\varphi_N}^T\varphi_N)^{-1}{\varphi_N}^Ty_N \quad \text{or} \quad \hat{\theta} = \rho_{\varphi_N\varphi_N}^{-1}\rho_{\varphi_N y_N}$$

ρ: covariance matrix

Parameter estimation

$\hat{\theta} = (\varphi_N{}^T \varphi_N)^{-1} \varphi_N{}^T y_N$

ARX example

- For a simple ARX model

$$y(k) = a_1 y(k-1) + b_1 u(k) + e(k)$$

- The parameter set

$$\theta = \begin{bmatrix} a \\ b \end{bmatrix} \qquad \varphi = \begin{bmatrix} y(k-1) \\ u(k) \end{bmatrix} \qquad \varphi_N = \begin{bmatrix} \varphi^T(1) \\ \vdots \\ \varphi^T(N) \end{bmatrix} = \begin{bmatrix} y(-1) & u(0) \\ \vdots & \vdots \\ y(-N) & u(-N+1) \end{bmatrix}$$

$$\varphi_N^T = \begin{bmatrix} y(-1) & \cdots & y(-N) \\ u(0) & \cdots & u(-N+1) \end{bmatrix} \quad \varphi_N^T \varphi_N = \begin{bmatrix} y(-1) & \dots & y(-N) \\ u(0) & \dots & u(-N+1) \end{bmatrix} \begin{bmatrix} y(-1) & u(0) \\ \vdots & \vdots \\ y(-N) & u(-N+1) \end{bmatrix}$$

- The estimated parameter is

$$\hat{\theta} = \underbrace{\begin{bmatrix} \sum_{k=1}^{N} y^2(-k) & \sum_{k=1}^{N} y(-k)u(-k+1) \\ \sum_{k=1}^{N} u(-k)y(-k+1) & \sum_{k=1}^{N} u^2(-k+1) \end{bmatrix}^{-1}}_{\text{Covariance matrix of } y \text{ and } u} \begin{bmatrix} \sum_{k=1}^{N} y(-k)y(-k+1) \\ \sum_{k=1}^{N} u(-k+1)y(-k+1) \end{bmatrix}$$

Parameter estimation

Minimizing the prediction error

- For general case, we calculate the prediction error as usual:
$$\varepsilon(k,\theta) = y(k) - \hat{y}(k|\theta)$$
- With N data samples, the estimation of total error, or loss is:
$$V_N(\theta) = \frac{1}{N}\sum_{k=1}^{N} \varepsilon^2(k,\theta)$$
- The goal is to find
$$\hat{\theta}_N = \arg\min_{\theta} V_N(\theta)$$
- BTW, we may use any arbitrary positive, scalar-valued function $\ell(\varepsilon)$ as a measure and minimize
$$V_N(\theta) = \frac{1}{N}\sum_{k=1}^{N} \ell\big(\varepsilon(k,\theta)\big)$$

Parameter estimation

Model order and prediction error

- If the prediction error has to be compared with model data
  - A larger model will do a better job
    - It also models the particular disturbance
    - But it will do worse with new data

- Alternative methods to estimate the prediction error penalizing model order p, of the general form:

$$\min_{p,\theta} f(p,N) \sum_{k=1}^{N} \varepsilon^2(k,\theta)$$

  - Akaike's information criterion (AIC)

$$\min_{p,\theta} \left(1 + \frac{2p}{N}\right) \sum_{k=1}^{N} \varepsilon^2(k,\theta)$$

  - Final prediction error (FPE)

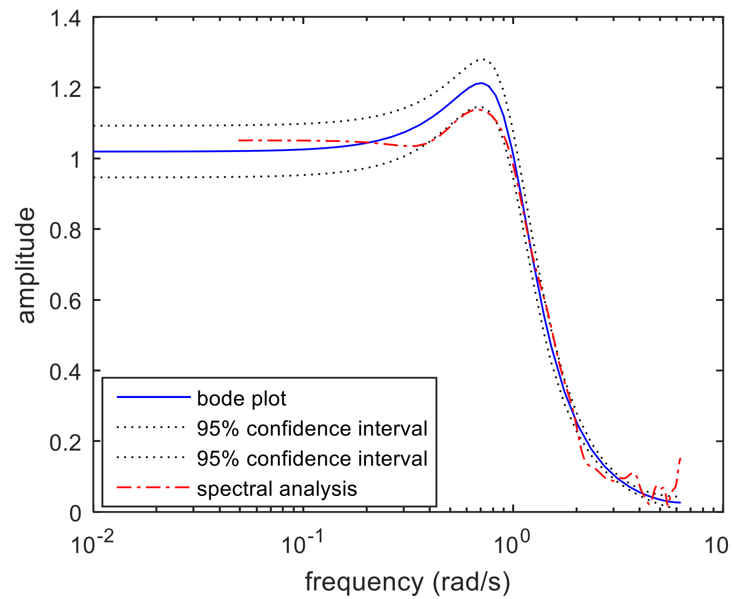
$$\min_{p,\theta} \left(\frac{1 + p/N}{1 - p/N} \frac{1}{N}\right) \sum_{k=1}^{N} \varepsilon^2(k,\theta)$$

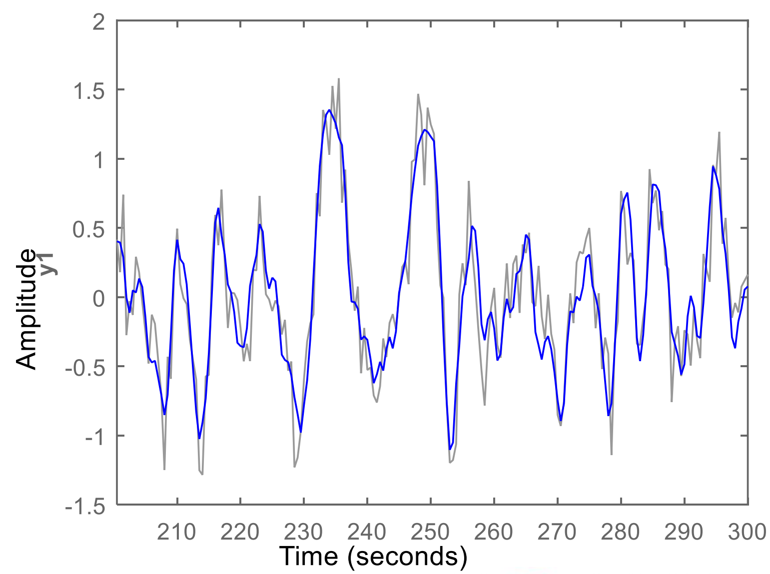
  - Rissanen's minimal description length

$$\min_{p,\theta} \left(1 + \frac{2p}{N} \log N\right) \sum_{k=1}^{N} \varepsilon^2(k,\theta)$$

Model validation

- Check if a model can be accepted for the intended use
  - Closely related to model quality
- Model quality
  - Stability
    - Input-output properties with different measurements
      - Bode diagram, simulation
    - Compare bode plot of the model vs. spectral analysis
      - Except for closed-loop system
  - Ability to reproduce system behavior
    - Compare simulated output vs. new measured data
    - Residual analysis





Model structure and order selection

- Residual analysis
- Zeros and Poles cancellation
- Variance of estimated parameter

Residual analysis: whiteness and independence tests

- The autocorrelation of

$$\varepsilon(k) = y(k) - \hat{y}\left(k \middle| \hat{\theta}_N\right)$$

$$\hat{\rho}_{\varepsilon\varepsilon}(\tau) = \frac{1}{N}\sum_{k=1}^{N} \varepsilon(k)\varepsilon(k-\tau)$$

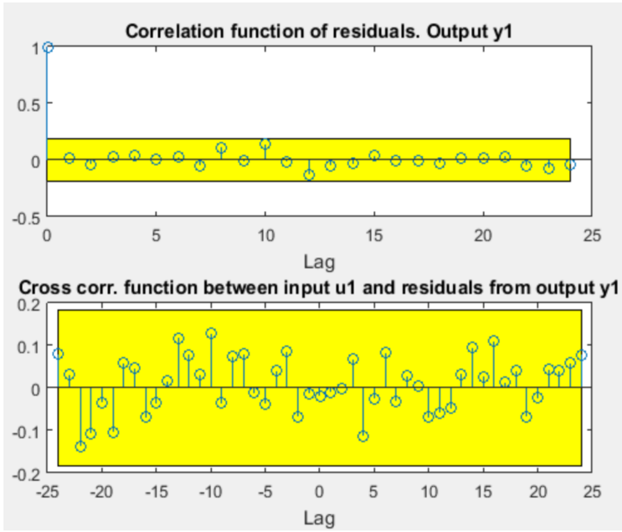
- Should lie within a confidence interval around zero
  - Large components indicates unmodelled disturbance

- The cross-correlation between $\varepsilon(k)$ and $u(k)$

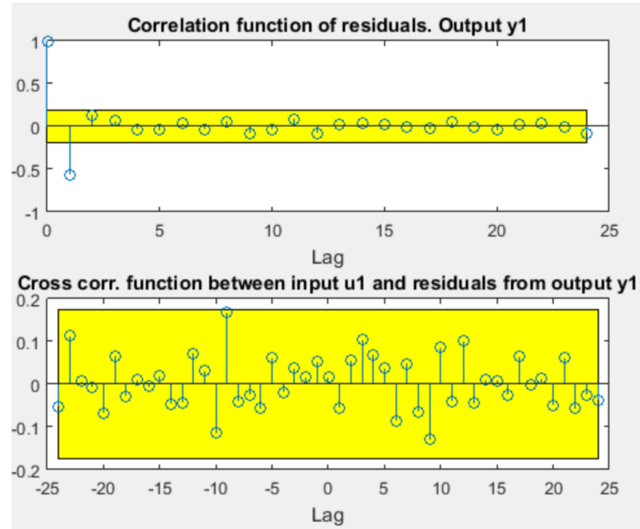
$$\hat{\rho}_{\varepsilon u}(\tau) = \frac{1}{N}\sum_{k=1}^{N} \varepsilon(k+\tau)u(k)$$

- Should lie within a confidence interval around zero
  - Large components indicate deficiency in system model
    - Model order problem

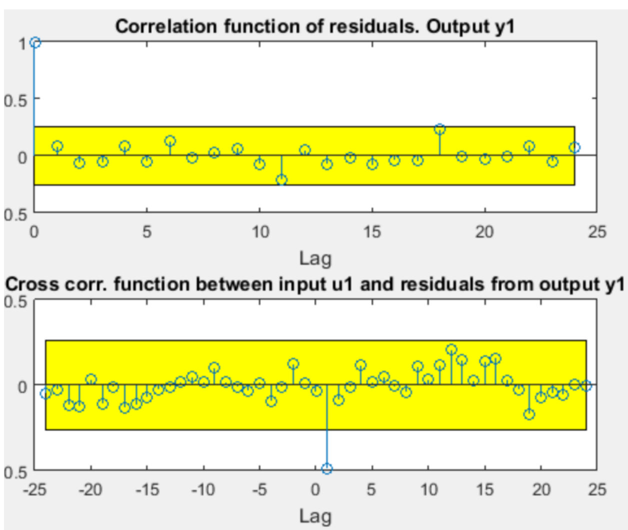
Examples of residual analysis



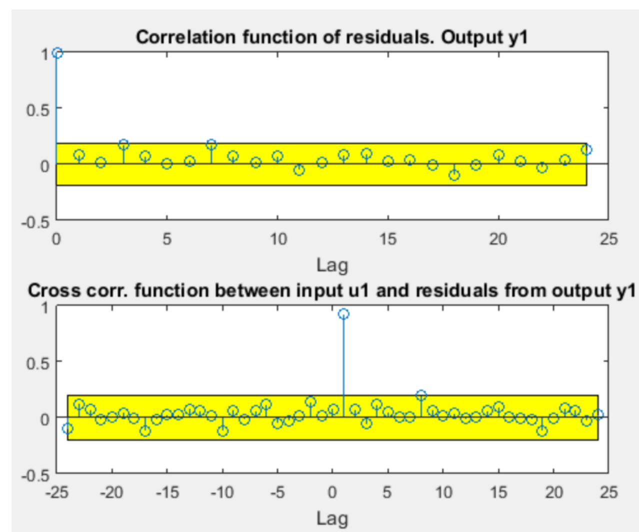
Model is acceptable



Error in disturbance model

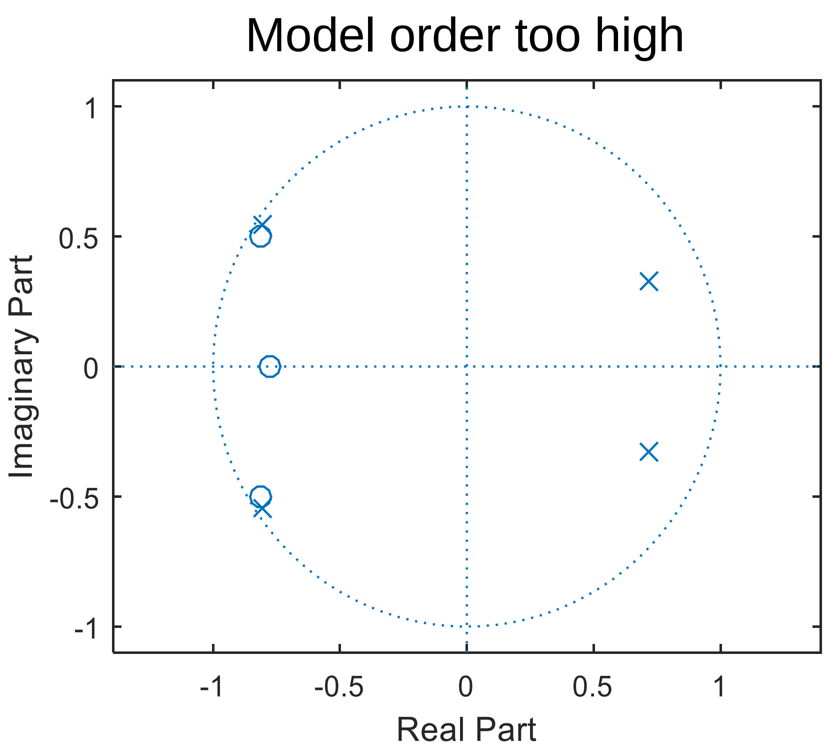


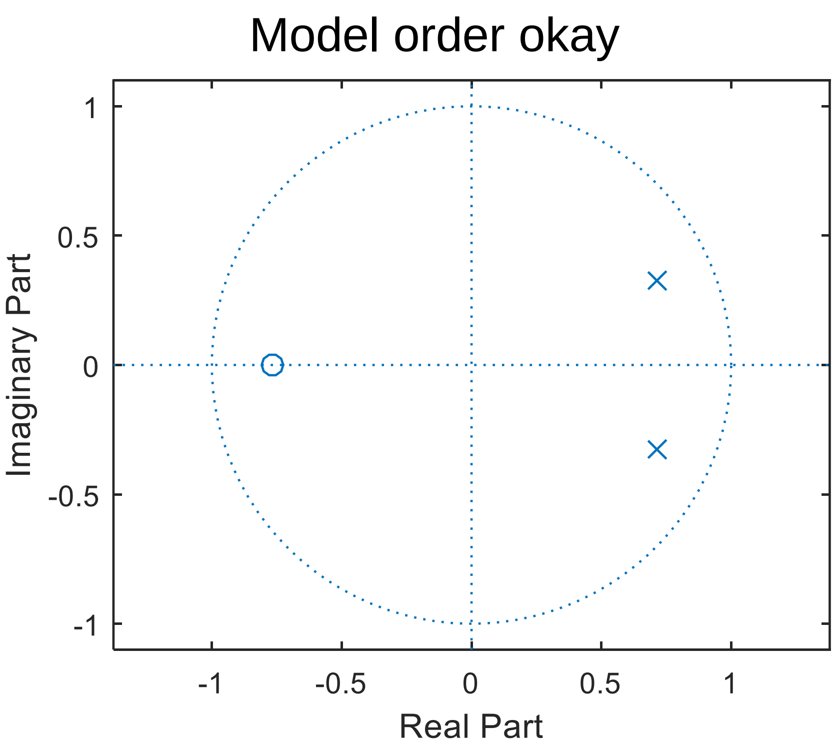
Error in system model



Delay too long

Zeros and Poles





Aalto University
School of Electrical
Engineering

Variance of estimated parameter

- The variance of the model parameter

$$P_N = E[\,(\theta - \theta_0)(\theta - \theta_0)^T] \approx \frac{1}{N}\lambda \bar{R}^{-1}$$

  - λ is the variance of the disturbance
  - $\bar{R} = E[\psi(k,\theta_0)\psi^T(k,\theta_0)]$
  - $\psi(k,\theta_0) = \frac{d}{d\theta}\hat{y}(k,\theta)$

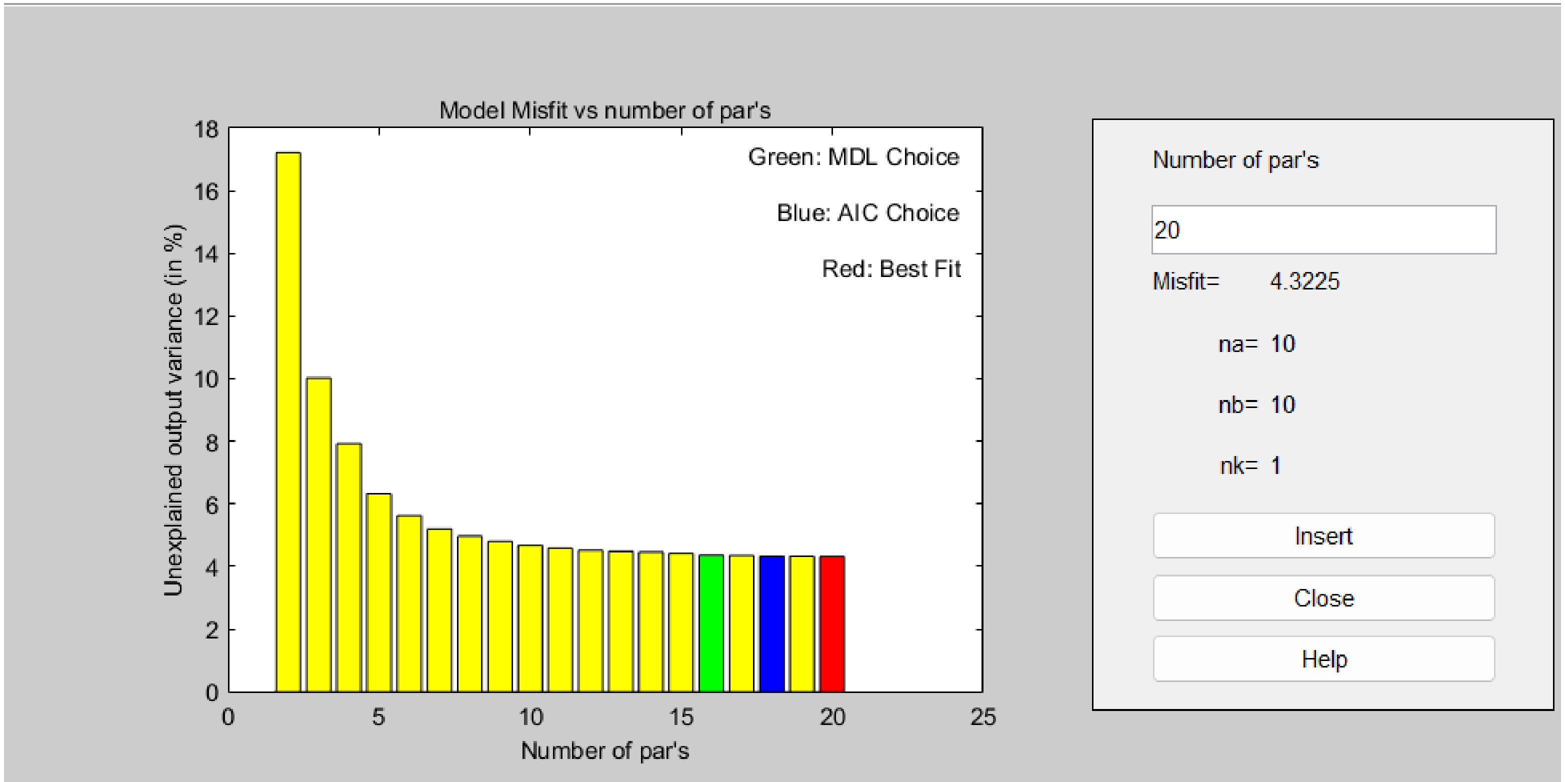
The variance should be at most 25% of the parameter value

- Example 1
  - $A(q) = 1 - 1.397\ (\pm 0.02608)q^{-1} + 0.5866\ (\pm 0.01946)q^{-2}$
  - $B(q) = 0.2026\ (\pm 0.01475)q^{-2} - 0.02881\ (\pm 0.01828)q^{-3}$
  - $C(q) = 1 - 0.9909\ (\pm 0.1401)q^{-1} + 0.2294\ (\pm 0.1311)q^{-2}$

- Example 2
  - $A(q) = 1 - 1.425\ (\pm 0.01208)q^{-1} + 0.6122\ (\pm 0.01146)q^{-2}$
  - $B(q) = 0.1113\ (\pm 0.002952)q^{-1} + 0.08808\ (\pm 0.003689)q^{-2}$
  - $C(q) = 1 - 0.3811\ (\pm 0.04841)q^{-1}$

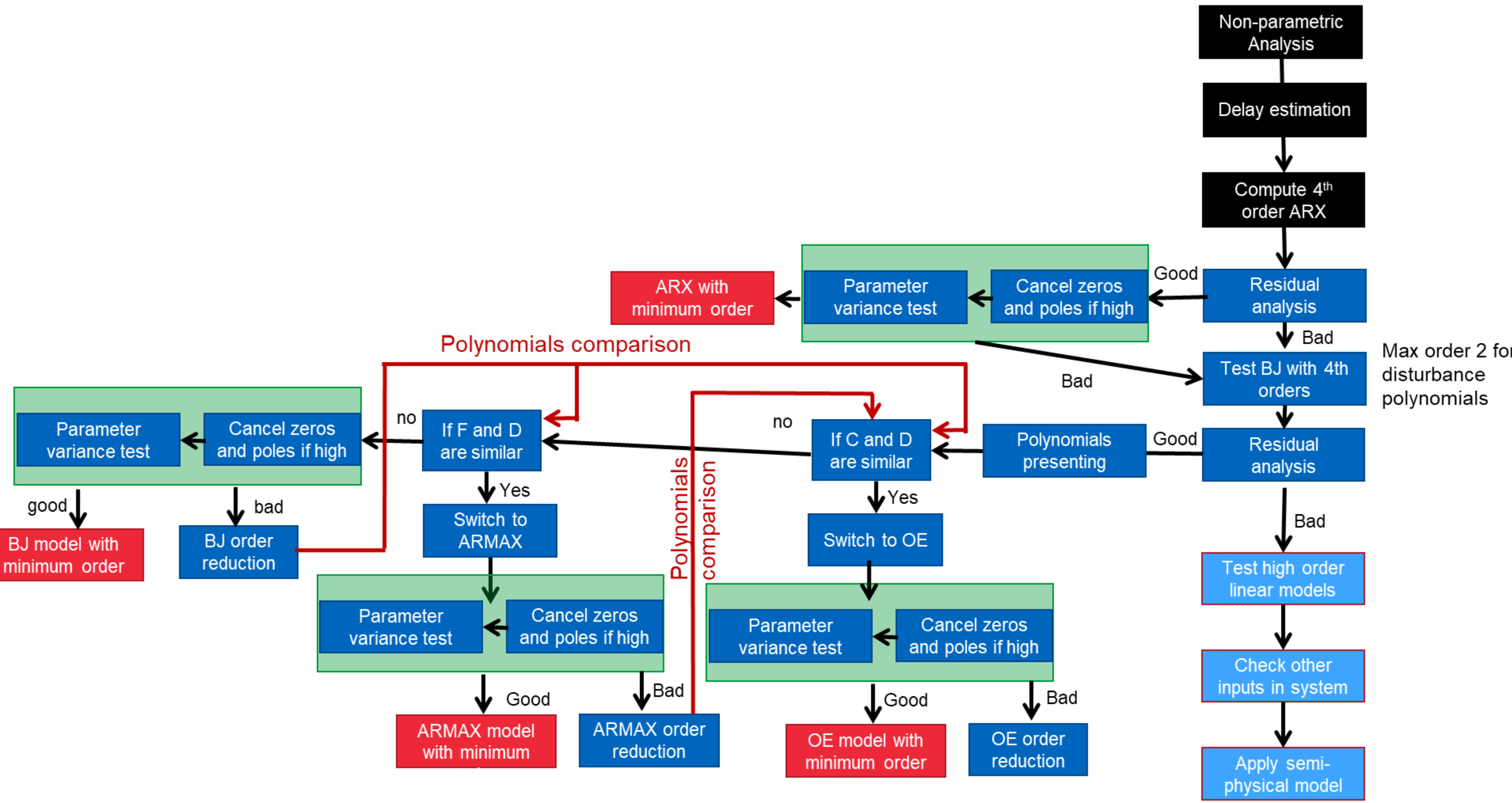
Delay estimation

Test ARX with different orders




```
data1 = iddata(y1,u1,1)
nk = delayest(data1)
```

Model structure and order selection strategy



Summary

- System identification is an iterative procedure in multiple steps
  - Experiment design
    - Preliminary experiments exams basic system behavior
    - Experiments test should excite the system
    - Take care of frequency band and sampling
  - Preprocess the data before identification
    - Remove mean, trends, and outlier
    - Filter if needed
  - Select the model structure
    - Parameters variance,
    - zeros and poles,
    - Residual analysis
  - Parameters estimation
    - Prediction error
  - Model validation
    - Cross-validation
    - Residual analysis