Welcome to...

MUDE

Modelling, Uncertainty, and Data for Engineers

WEEK 7 !

Modelling, Uncertainty and Data for Engineers (MUDE)

Week 1.7-1.8 : Sensing and Observation Theory

Sandra Verhagen

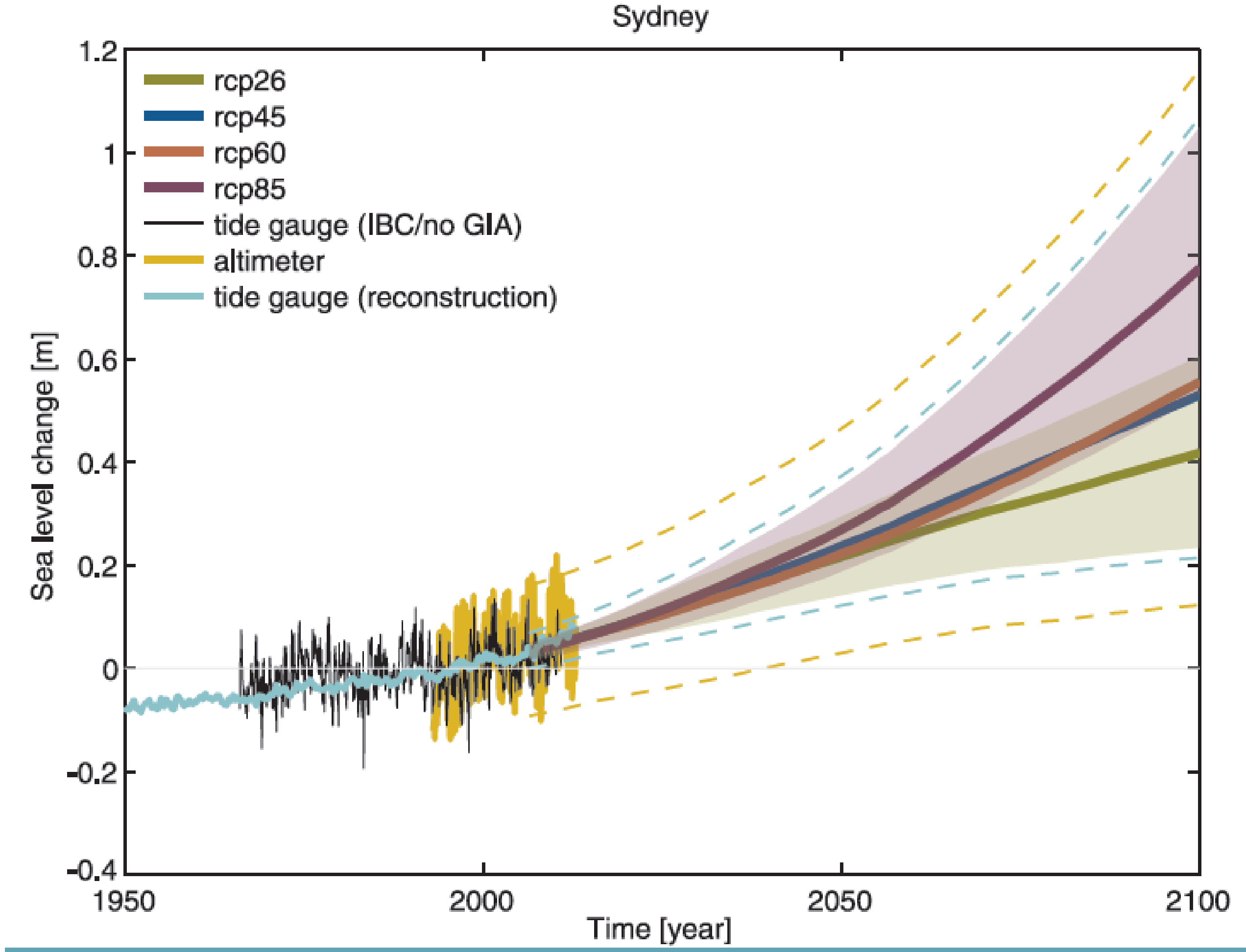
TU Delft

Sensing and observation theory

- Science and engineering: need observations!
- Observations → parameters of interest?
- Estimation results: interpretation & uncertainty

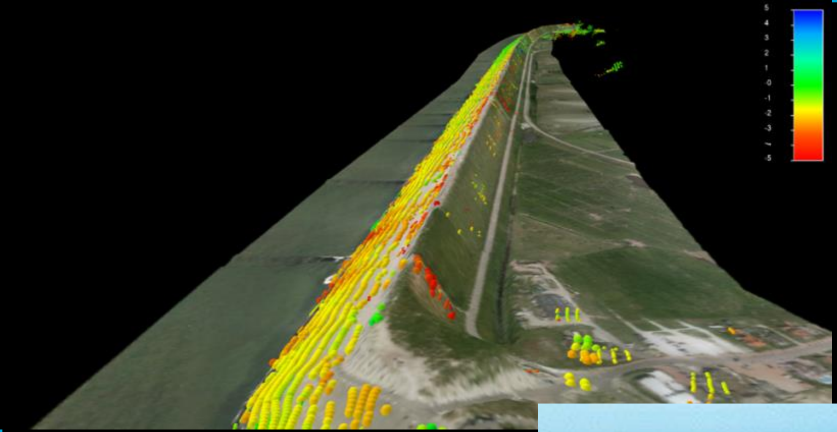
→ Input for other engineers, decision makers, …

https://research.csiro.au/slrwavescoast/sea-level/future-sea-level-changes/



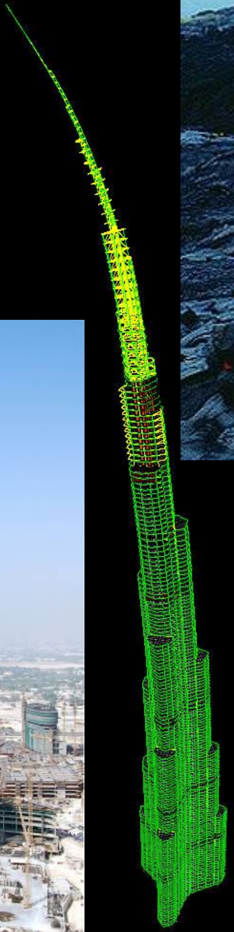
Do we need higher dikes?

Monitoring and Sensing: why?









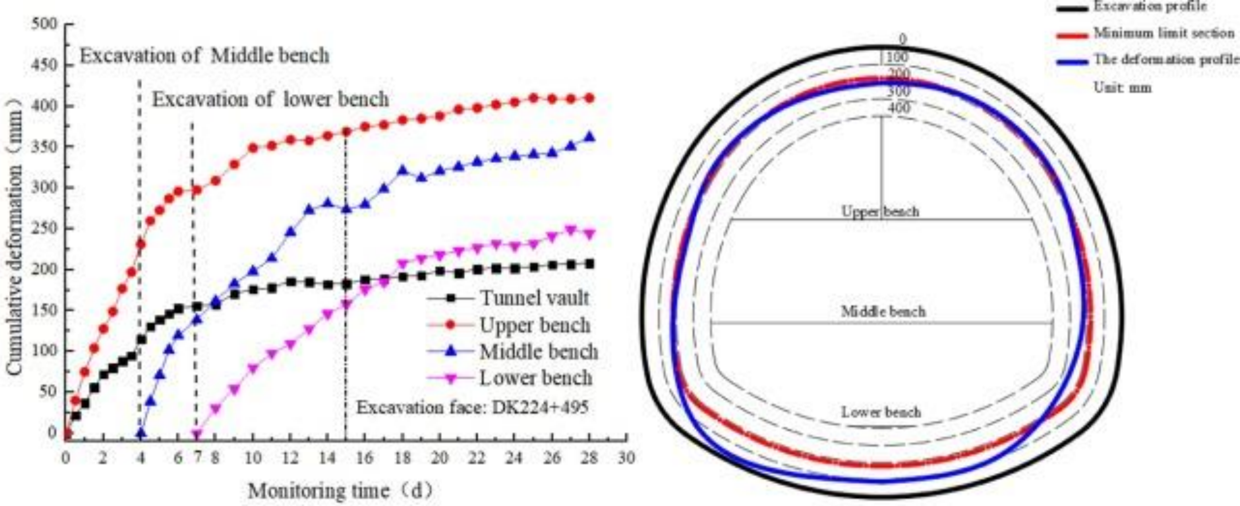
What sensor / observation types are used in your discipline?

Sensor/observation types

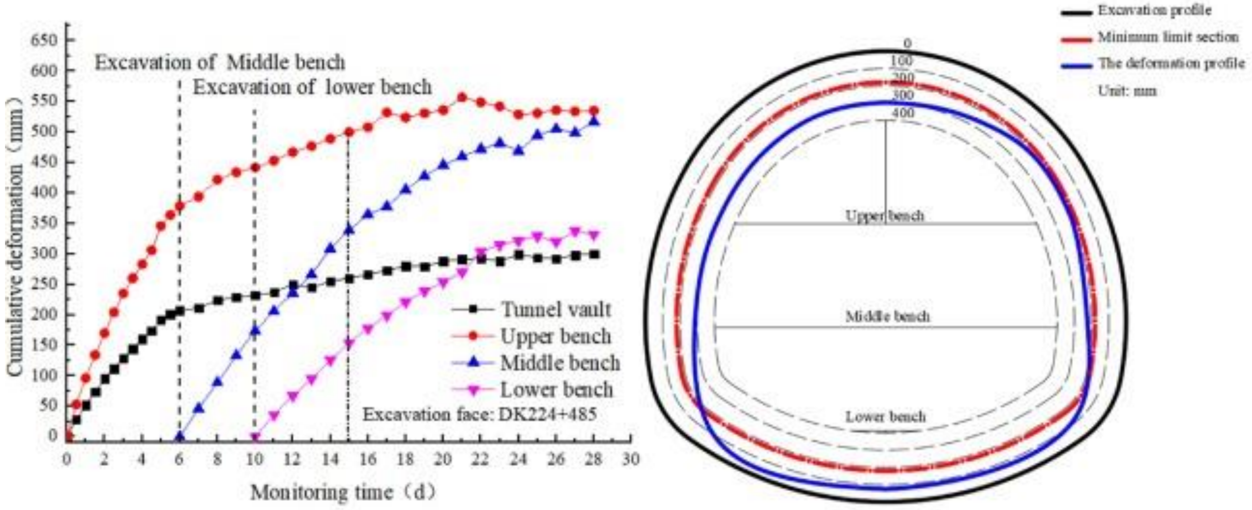
- camera: visible, IR, UV, hyperspectral
- radar
- radio signals
- rain gauges
- tide gauges
- stress / strain sensors
- acoustic sensors
- accelerometers
- gyroscopes
- temperature
- pressure
- ….

Sensing and observation theory: applications

- Sea level rise
- Subsidence / uplift
- Air quality modelling
- Settlement of soils
- Tunnel deformation
- Bridge motions
- Traffic flow rate
- Water vapor content
- Ground water level



(a) DK224+530 Section monitoring data and deformation profile



(b) DK224+520 Section monitoring data and deformation profile

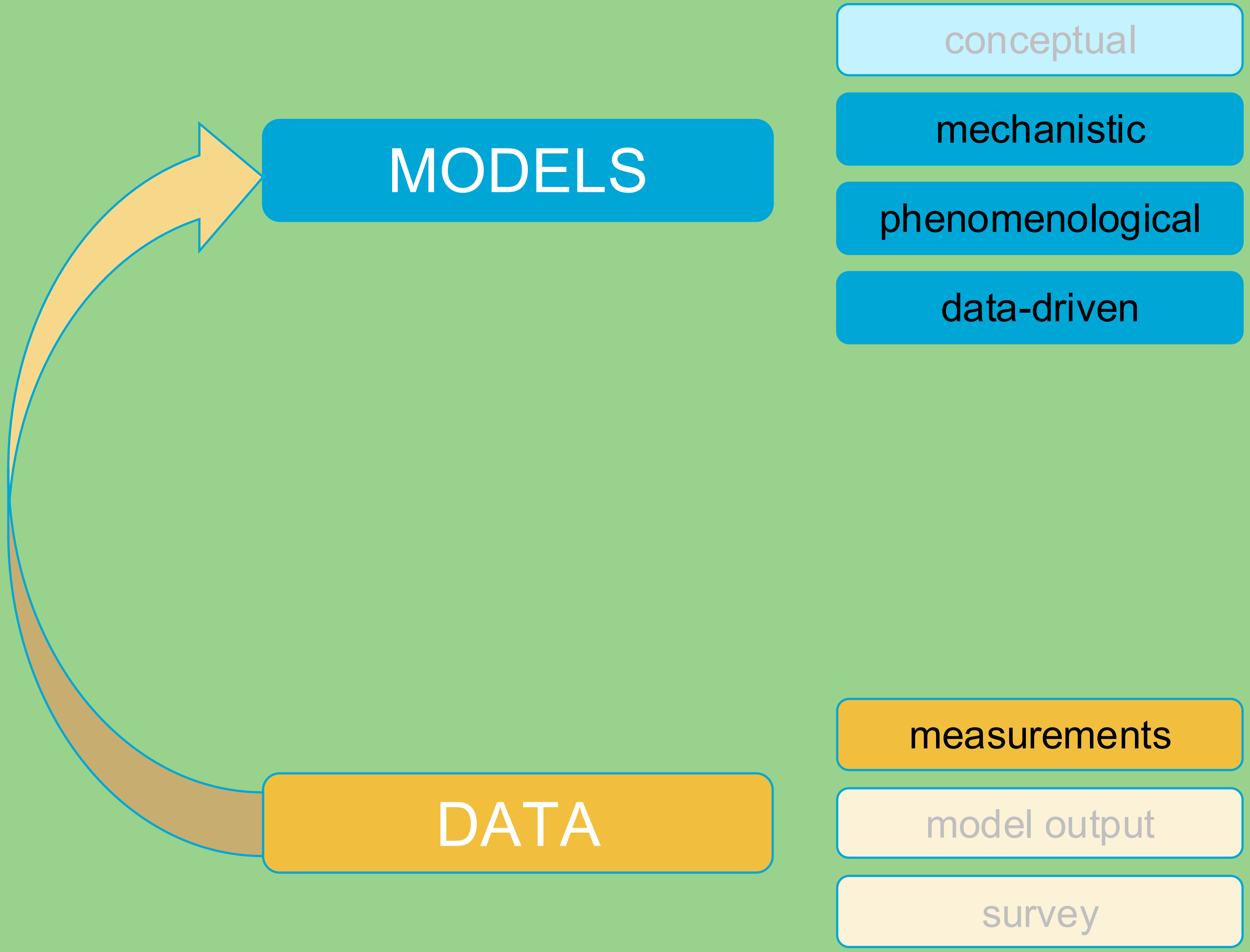
Input data Y → model & estimate parameters of interest x → Output data $\hat{X} = q(Y)$

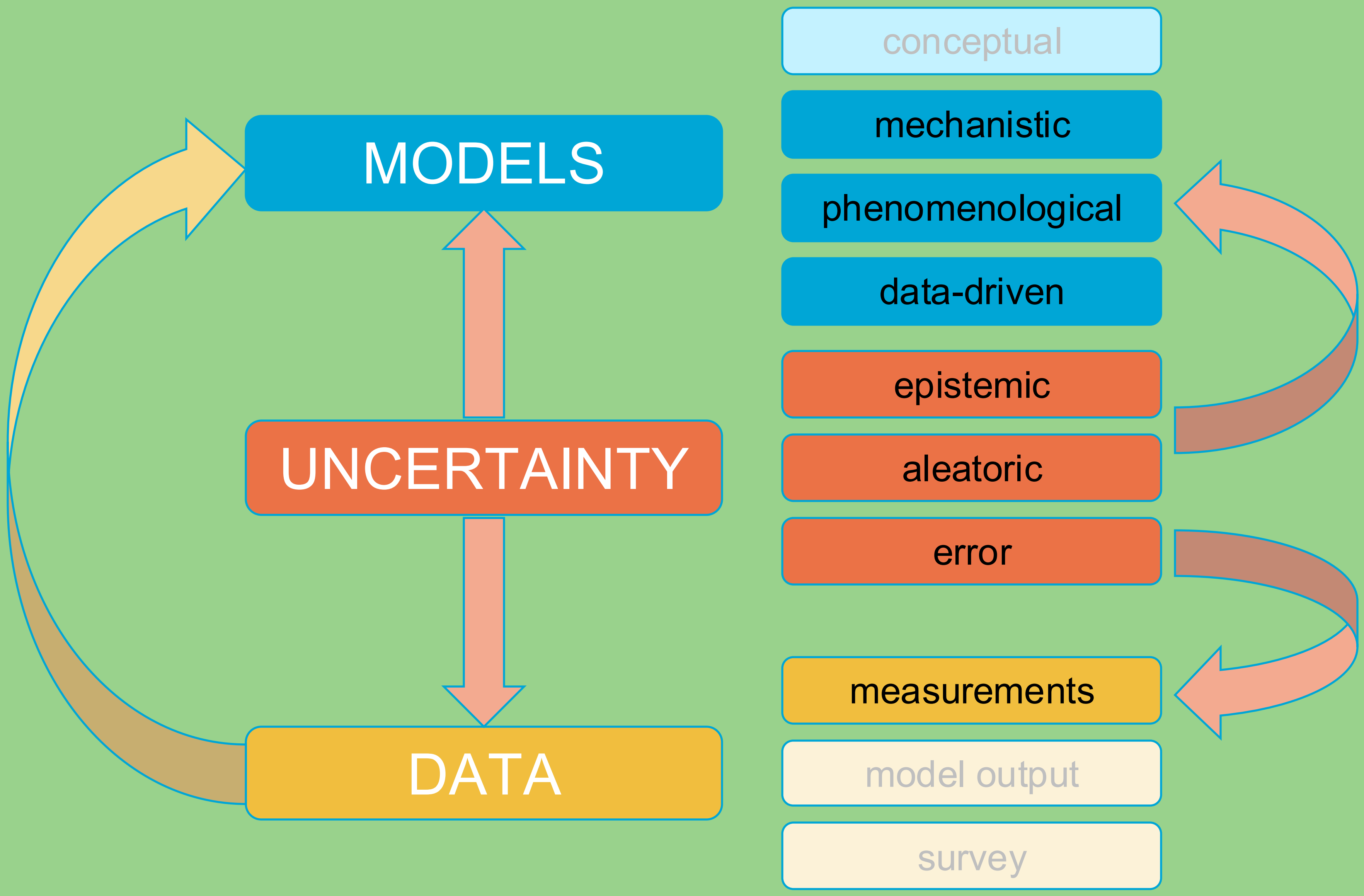
DATA

measurements

model output

survey


conceptual
mechanistic
phenomenological
data-driven
MODELS
DATA
measurements
model output
survey



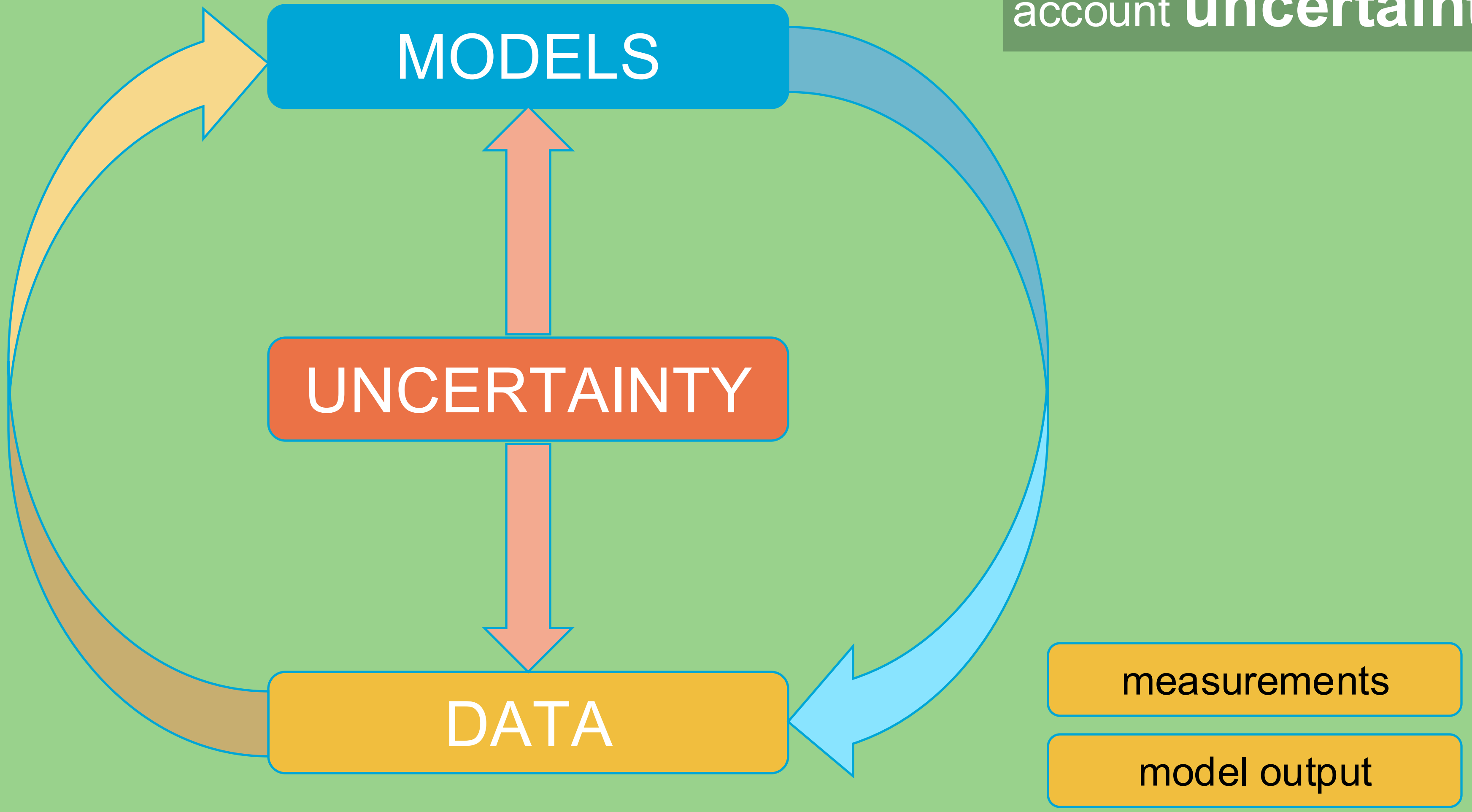
MODELS
UNCERTAINTY
DATA
conceptual
mechanistic
phenomenological
data-driven
epistemic
aleatoric
error
measurements
model output
survey

What sensor / observation types are used in your discipline?

Sensor/observation types

- camera: visible, IR, UV, hyperspectral
- radar
- radio signals
- rain gauges
- tide gauges
- stress / strain sensors
- acoustic sensors
- accelerometers
- gyroscopes
- temperature
- pressure
- ….

This week is about:
estimating model parameters using **data** (usually measurements), taking into account **uncertainty** in data and models



Ingredients

Input data Y → model & estimate parameters of interest x → Output data $\hat{X} = q(Y)$

You wil need …

- … a model to describe relation between Y and x
- … to select and apply an appropriate estimation method
- … to apply uncertainty propagation to assess the precision of $\hat{X}$
- … to apply tests to assess validity of your model

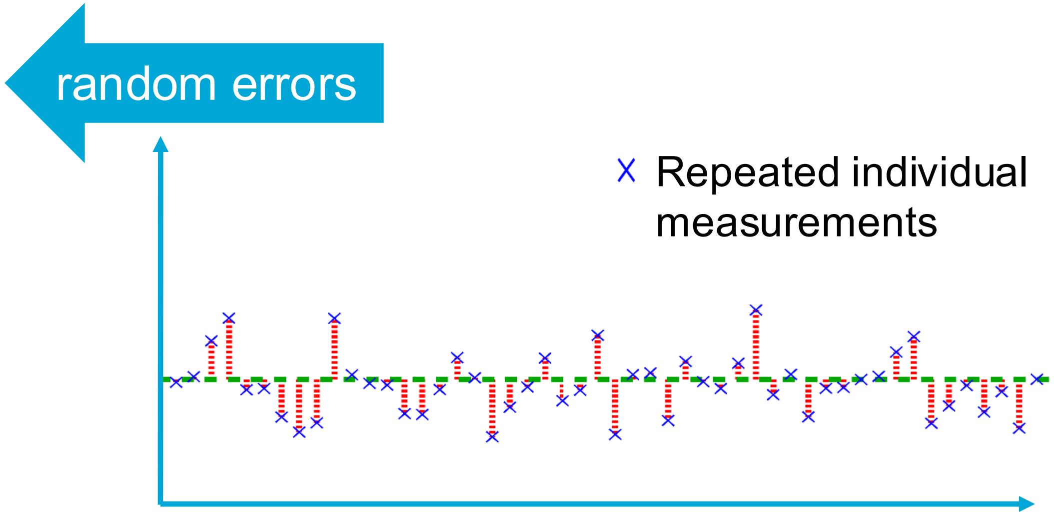
TU Delft

Ingredients

Input data Y → model & estimate parameters of interest x → Output data $\hat{X} = q(Y)$

You wil need …

- … a model to describe relation between Y and x
- … to select and apply an appropriate estimation method
- … to apply uncertainty propagation to assess the precision of $\hat{X}$
- … to apply tests to assess validity of your model

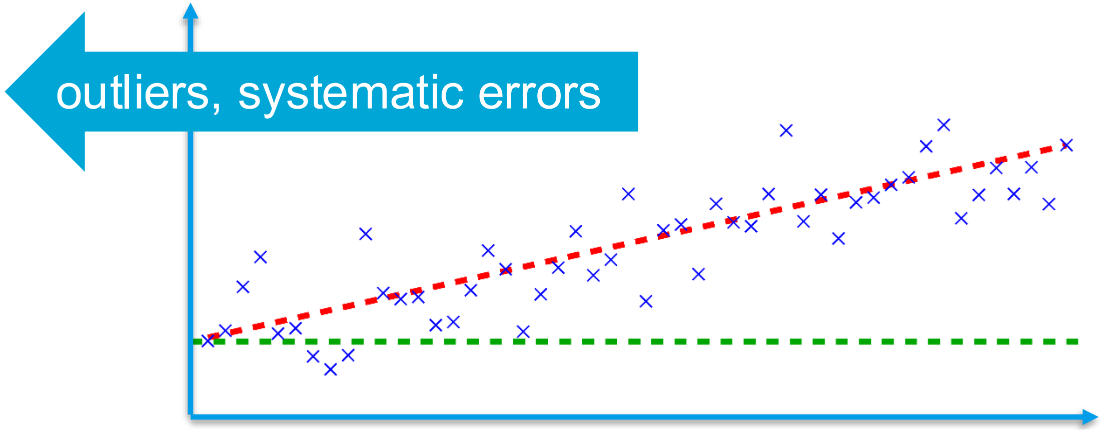


Ingredients

Input data Y → estimate parameters of interest x → Output data $\hat{X} = q(Y)$

You wil need …

- … a model to describe relation between Y and x
- … to select and apply an appropriate estimation method
- … to apply uncertainty propagation to assess the precision of $\hat{X}$
- … to apply tests to assess validity of your model
    - to account for errors in data



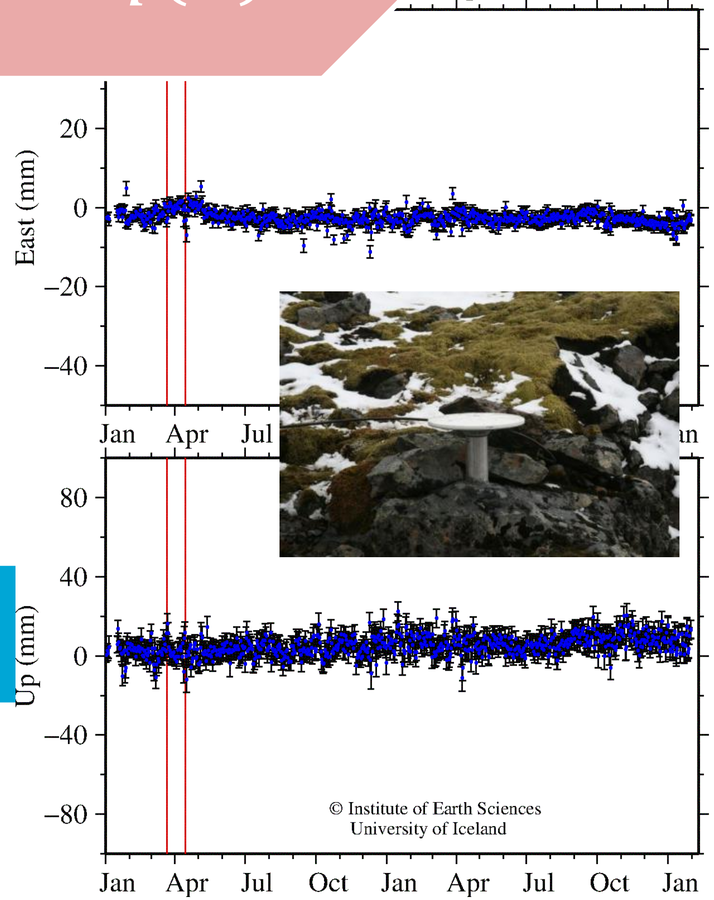
Ingredients

Input data Y → estimate parameters of interest x → Output data $\hat{X} = q(Y)$

You wil need …

- … a model to describe relation between Y and x
- … to select and apply an appropriate estimation method
- … to apply uncertainty propagation to assess the precision of $\hat{X}$
- … to apply tests to assess validity of our model
  - to account for errors in data
  - to choose best model from different candidates

example:
change detection



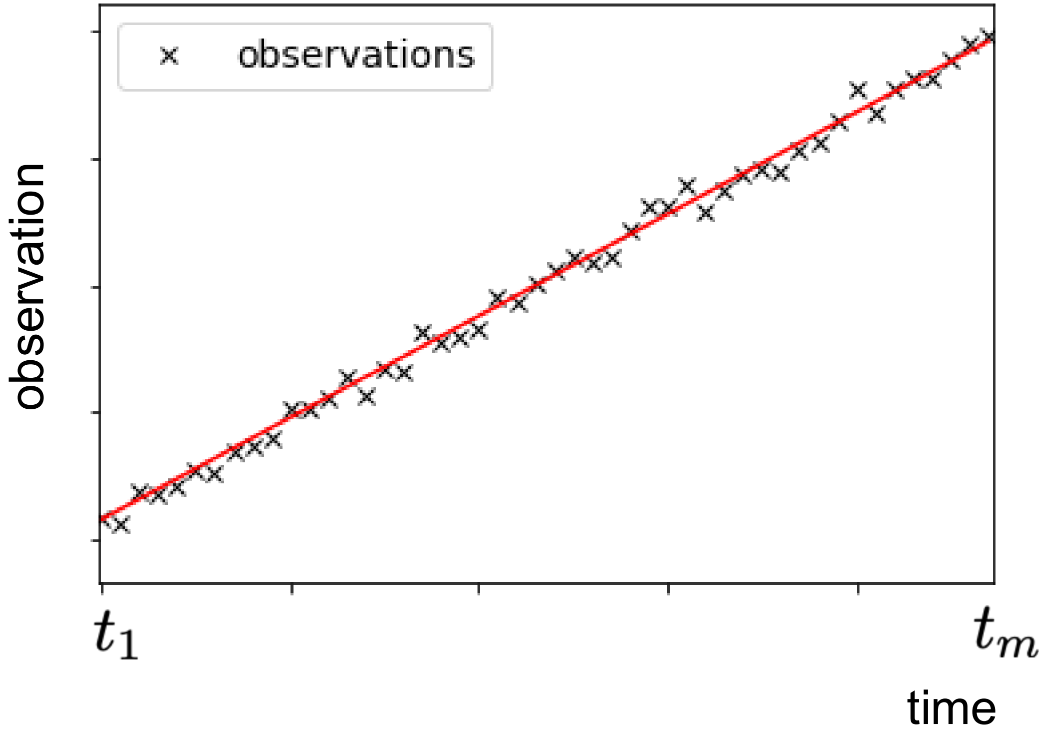
TU Delft

Examples

Linear trend model:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_m \end{bmatrix}$$

$$= \mathrm{Ax} + \epsilon$$



Unknowns (model parameters):

x_1 initial value at $t = 0$

x_2 slope

Model formulation

- Observable Y : stochastic quantity (due to random errors)
  → an observable ("to be observed quantity") has a certain probability distribution
- Observation vector y : realization of Y
  → the measured value(s)
- Parameter vector x : deterministic, but unknown
- Random errors ϵ : stochastic with $\epsilon \sim N(0, \Sigma_\epsilon)$
- Functional model (linear case) : $Y = \underset{m \times n}{\mathrm{A}} \cdot \mathrm{x} + \epsilon$ or $\mathbb{E}(Y) = \mathrm{A} \cdot \mathrm{x}$

What is the distribution of Y?

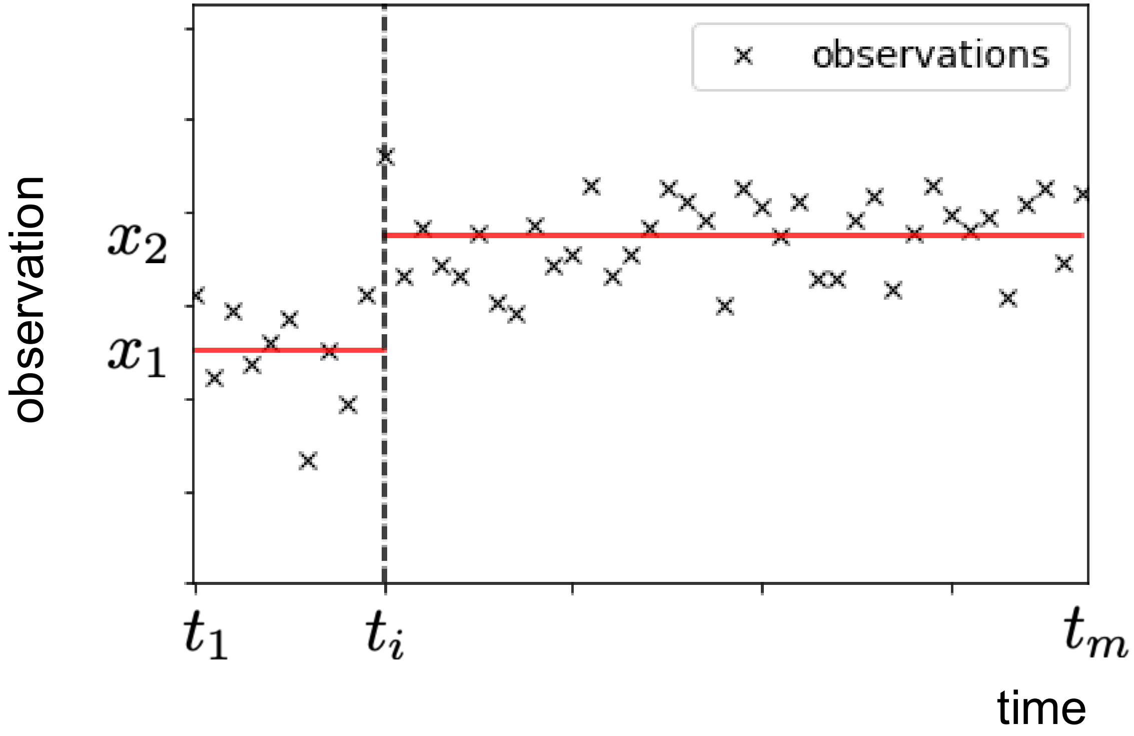
- Design matrix A : describes functional relationship between Y and x

TUDelft

Examples

Step model

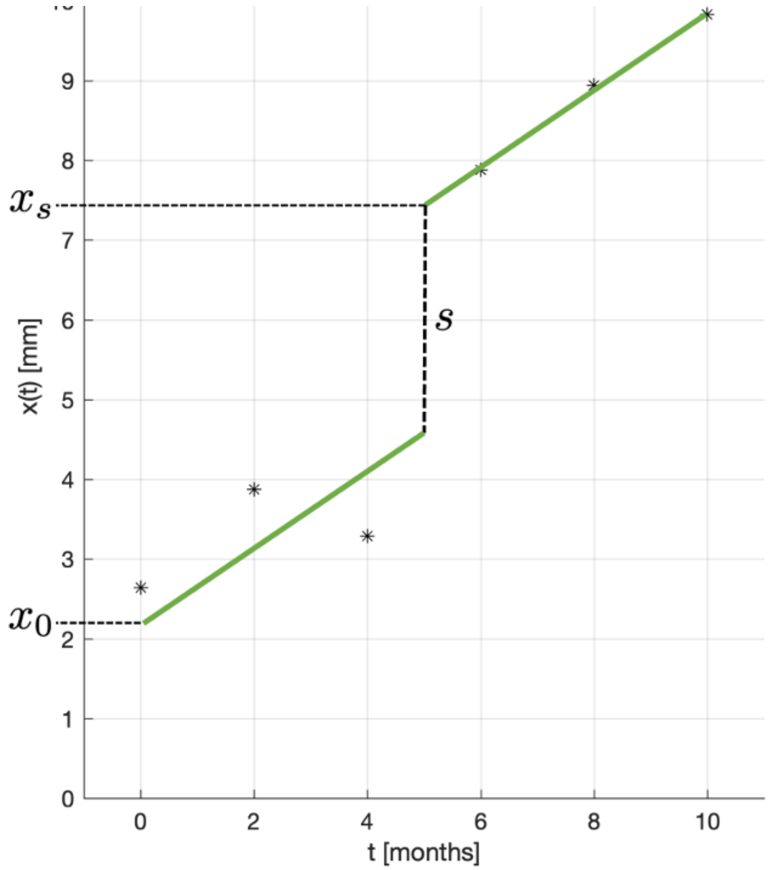
$$\mathbb{E}(\begin{bmatrix} Y_1 \\ \vdots \\ Y_{i-1} \\ Y_i \\ \vdots \\ Y_m \end{bmatrix}) = \underbrace{\begin{bmatrix} \\ \\ \\ \\ \\ \\ \end{bmatrix}}_{\mathrm{A}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\mathrm{x}}$$



Part 7: Sensing and Observation

The distance x between a fixed benchmark and a moving benchmark on a landslide is measured at times $t = 0, 2, 4, 6, 8, 10$ months. The observations are shown in the figure.

It is assumed that normally the distance is changing at a constant rate. It is known, however, that at $t = 5$ months there was a sudden slip of the landslide, causing an additional change in distance at that time.



Observations y collected,
we have a functional model A,
how to estimate x?

Observations y collected, we know A, how to estimate x?

for now we ignore the random errors

A linear system $\mathrm{y} = \underset{m \times n}{\mathrm{A}} \cdot \mathrm{x}$

We will consider overdetermined systems with $rank(\mathrm{A}) = n < m$

Hence we have more observations than unknowns

Redundancy = $m - n$

TU Delft

Example of overdetermined system with $rank(\mathrm{A}) = n$

$$\underbrace{\begin{bmatrix} 3 \\ 5 \\ 6 \end{bmatrix}}_{\mathrm{y}} = \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}}_{\mathrm{A}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\mathrm{x}}$$

→ no solution

$$\underbrace{\begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}}_{\mathrm{y}} = \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}}_{\mathrm{A}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\mathrm{x}}$$

→ $\hat{\mathrm{x}} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$

→ in case of perfect measurements, i.e., errors equal to 0

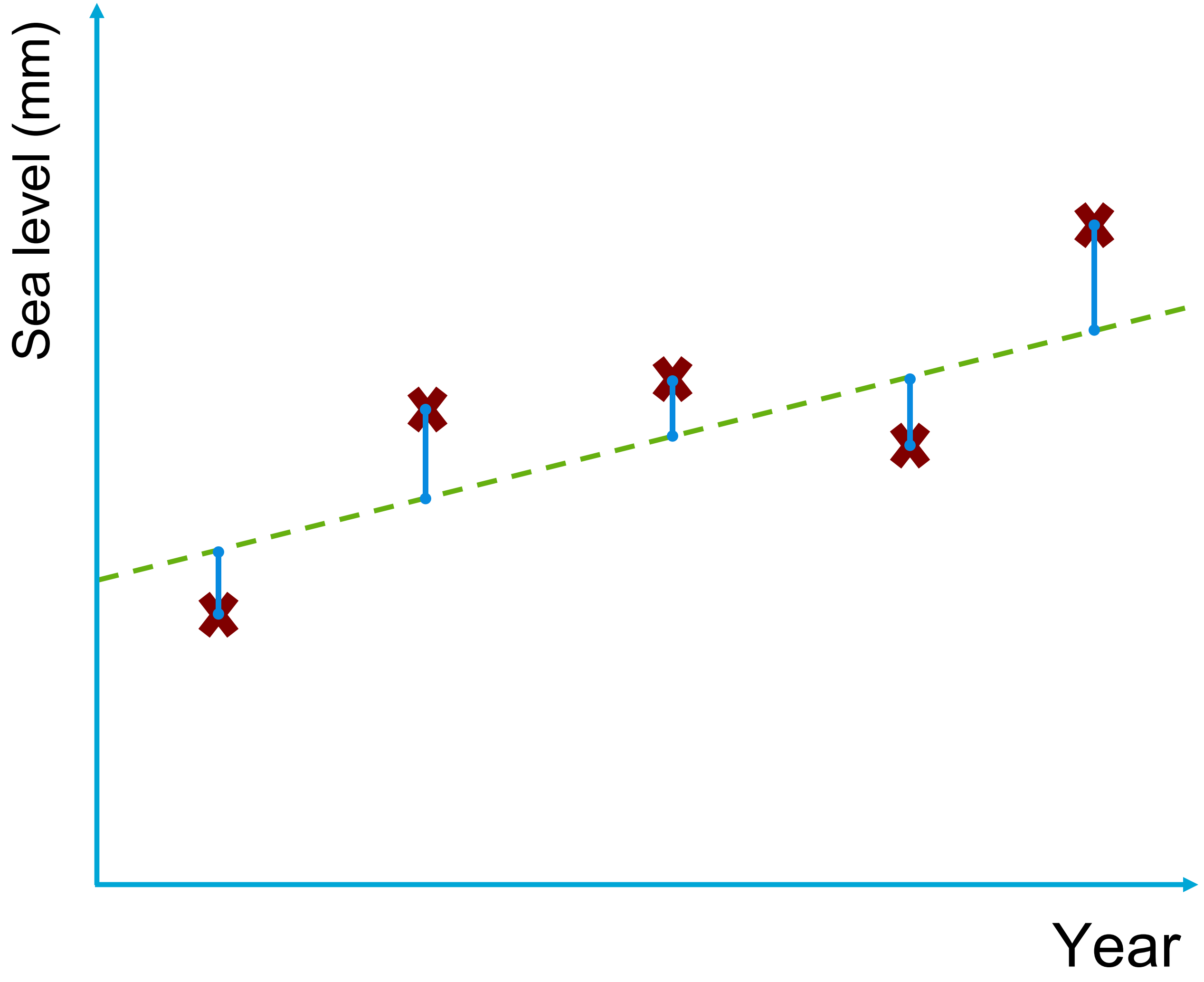
TU Delft

Overdetermined system

Account for random errors,
otherwise generally no solution

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \\ 1 & t_4 \\ 1 & t_5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix}$$

unknowns : 2 parameters + 5 errors
but only 5 observations…
many possible solutions



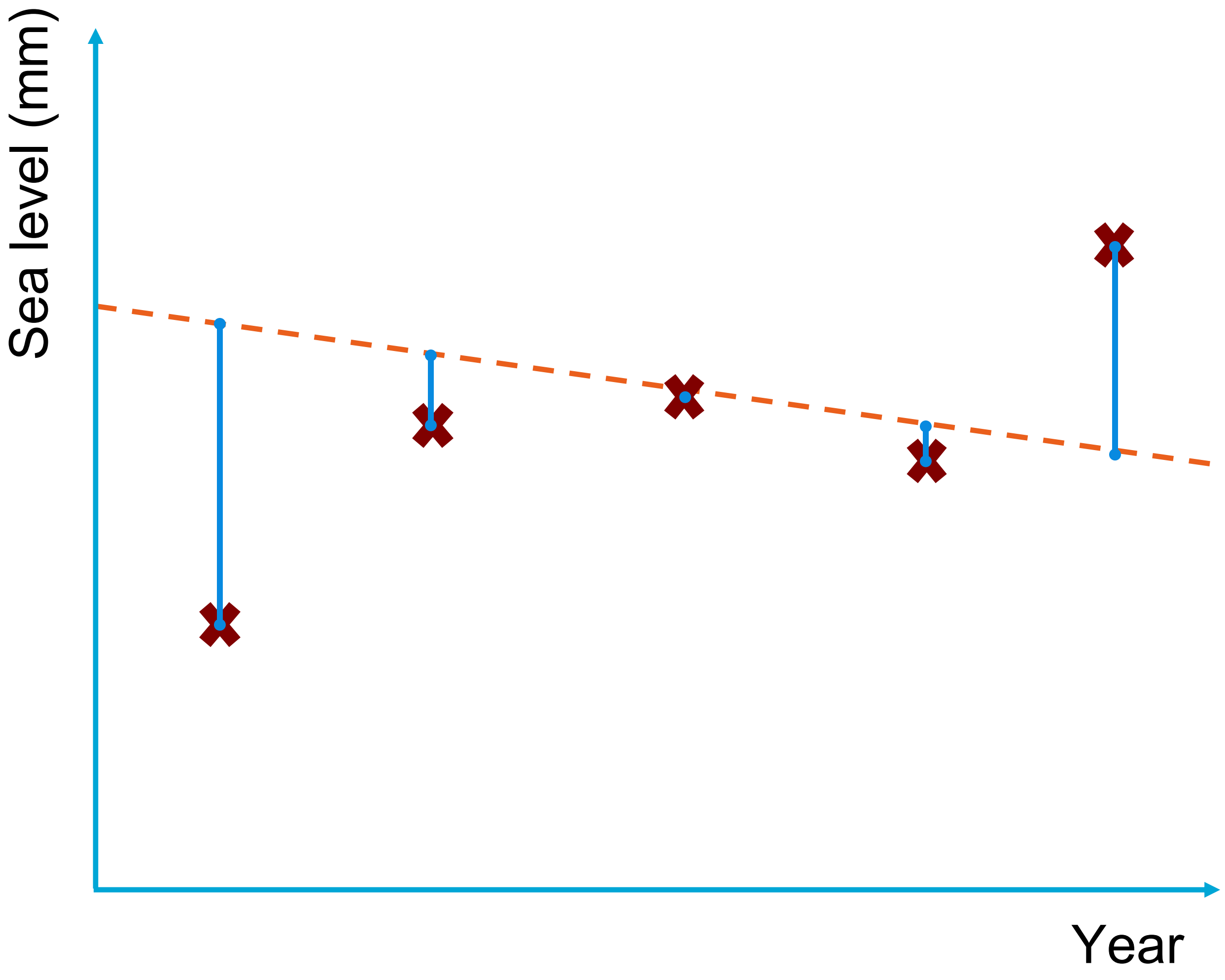
TUDelft

Overdetermined system

Account for random errors,
otherwise generally no solution

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \\ 1 & t_4 \\ 1 & t_5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix}$$

unknowns : 2 parameters + 5 errors
but only 5 observations…
many possible solutions



Least squares criterion?

Quiz: what is the least-squares criterion?

minimize the sum of the squared errors

TU Delft

Least-squares principle

- Linear model: $$\mathrm{y} = \mathrm{A}\mathrm{x} + \epsilon$$

- Objective: $$\min_{\mathrm{x}}(\epsilon^T \epsilon) = \min_{\mathrm{x}}(\mathrm{y} - \mathrm{A}\mathrm{x})^T(\mathrm{y} - \mathrm{A}\mathrm{x})$$

Minimize the sum of squared errors (i.e., optimization problem)

- Gradient (first-order partial derivatives) = 0
- Hessian (matrix with second-order partial derivatives) > 0

- Solution $$\hat{\mathrm{x}} = \left(\mathrm{A}^T\mathrm{A}\right)^{-1}\mathrm{A}^T \cdot \mathrm{y}$$

Least-squares solution

Functional model:

$$\mathrm{y} = \mathrm{A}\mathrm{x} + \epsilon$$

Least squares solution

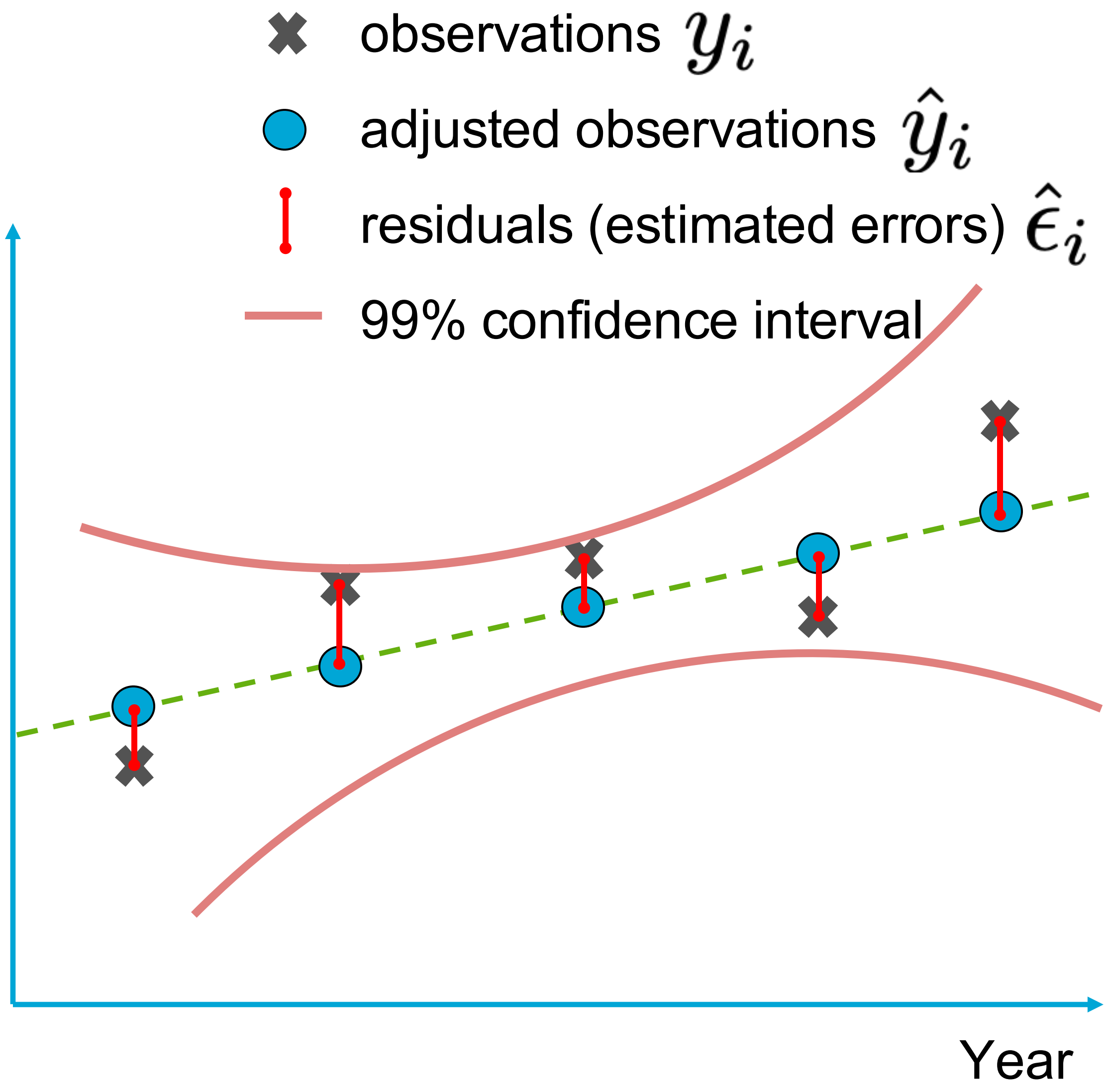
$$\hat{\mathrm{x}} = \left(\mathrm{A}^T \mathrm{A}\right)^{-1} \mathrm{A}^T \cdot \mathrm{y}$$

Adjusted (predicted) observations:

$$\hat{\mathrm{y}} = \mathrm{A}\hat{\mathrm{x}}$$

Residuals (estimated errors):

$$\hat{\epsilon} = \mathrm{y} - \hat{\mathrm{y}}$$



Properties of Least-Squares estimator

Modelling, Uncertainty and Data for Engineers

TU Delft

Estimate vs. Estimator

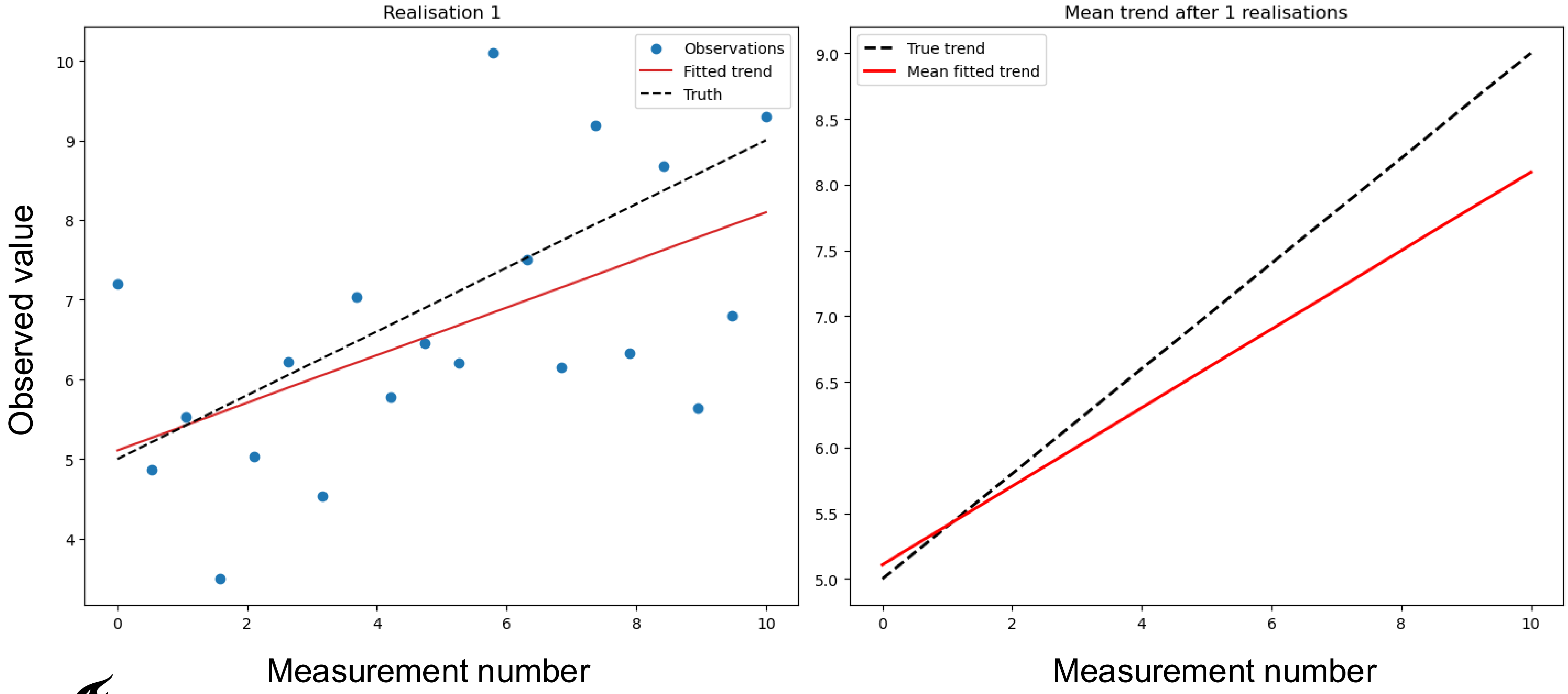
$$\hat{x} = (A^T A)^{-1} A^T \cdot y$$

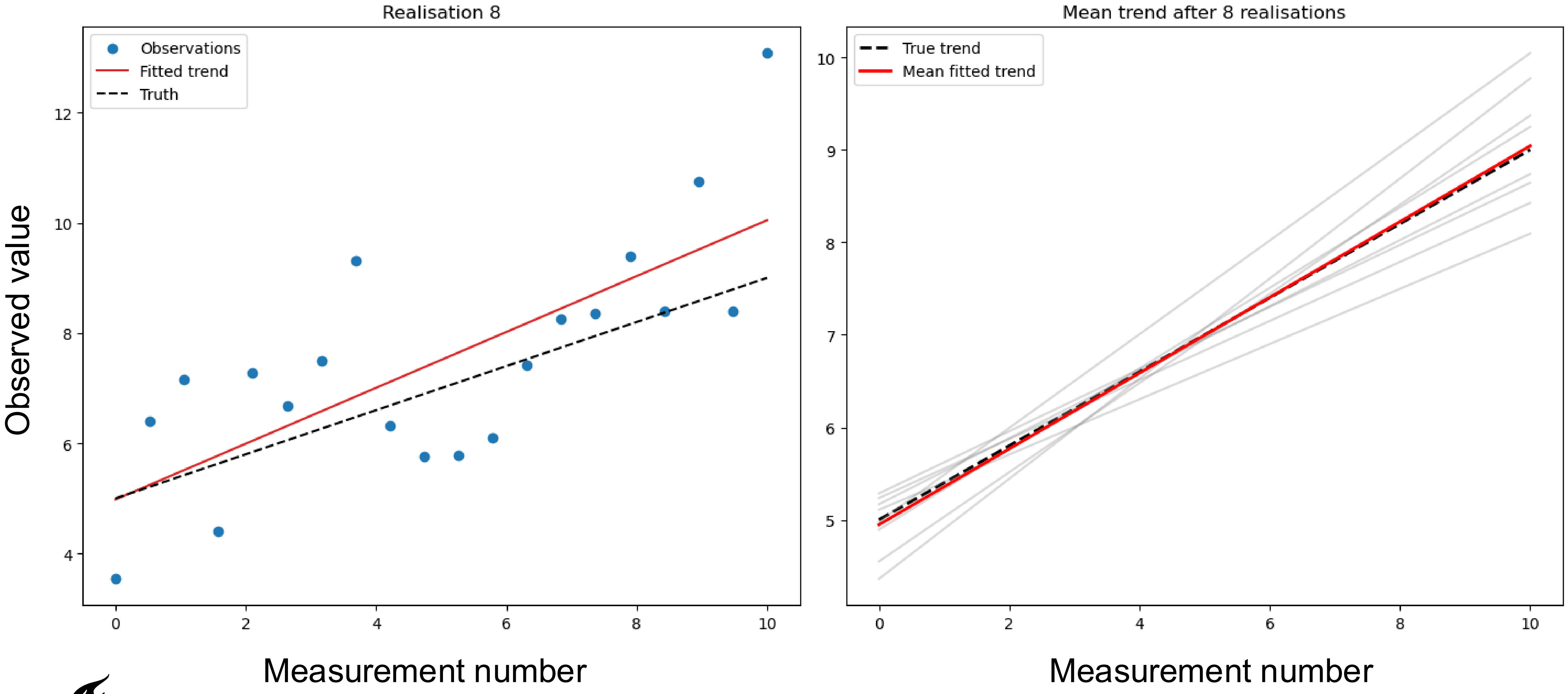
$$\hat{X} = (A^T A)^{-1} A \cdot Y$$

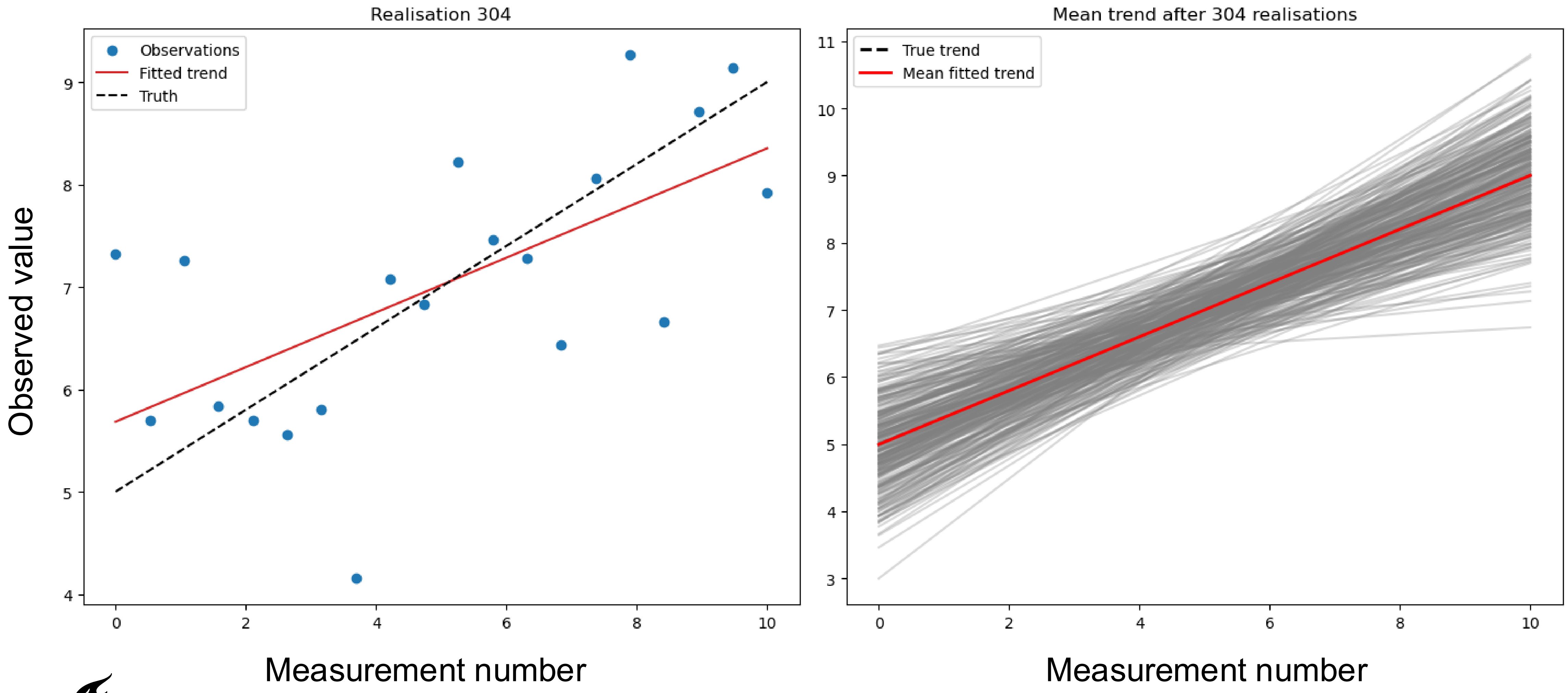
one realization of $\hat{X}$
(the estimated values)

one realization of Y
(the actually observed values)

- If you repeat the measurements, you will get a different realization (due to random errors) → estimated values will be different
- If you repeat many times → you will see distribution of estimated parameters and fitted model


Realisation 1
Observations
Fitted trend
Truth
Observed value
Measurement number
Mean trend after 1 realisations
True trend
Mean fitted trend
Measurement number


Realisation 8
Observations
Fitted trend
Truth
Observed value
Measurement number
Mean trend after 8 realisations
True trend
Mean fitted trend
Measurement number



Realisation 304
Observations
Fitted trend
Truth
Observed value
Measurement number
Mean trend after 304 realisations
True trend
Mean fitted trend
Measurement number

Estimate vs. Estimator

$$\hat{\mathrm{x}} = \left(\mathrm{A}^T\mathrm{A}\right)^{-1}\mathrm{A}^T \cdot \mathrm{y}$$

$$\hat{X} = (\mathrm{A}^T\mathrm{A})^{-1}\mathrm{A} \cdot Y$$

one realization of $\hat{X}$
(the estimated values)

one realization of Y
(the actually observed values)

- If you repeat the measurements, you will get a different realization (due to random errors) → estimated values will be different
- If you repeat many times → you will see distribution of estimated parameters and fitted model
- We can also determine the expectation and covariance matrix by applying the propagation laws you learned last week (Section 5.3)

TU Delft

From Section 5.3 Linear propagation of mean and covariance

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} + \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix} = \mathbf{AY} + \mathbf{c}$$

with known $\mathbb{E}(\mathbf{Y})$ and covariance matrix $\mathbf{\Sigma}_Y$, and $\mathbf{c}$ a vector with deterministic variables.

The linear propagation laws of the mean and covariance matrix are given by

$$\mathbb{E}(\mathbf{X}) = \mathbf{A}\mathbb{E}(\mathbf{Y}) + \mathbf{c}$$

$$\mathbf{\Sigma}_X = \mathbf{A}\mathbf{\Sigma}_Y\mathbf{A}^T$$

Applying the propagation laws...

$$\hat{X} = \underbrace{(\mathrm{A}^T\mathrm{A})^{-1}\mathrm{A}}_{\mathrm{L}^T} \cdot Y$$

in Section 6.10 Notation and formulas

Linear propagation laws if $\hat{X} = \mathrm{L}^{\mathrm{T}}Y$

Propagation law of the ...	Formula
... mean	$\mathbb{E}(\hat{X}) = \mathrm{L}^{\mathrm{T}}\mathbb{E}(Y)$
... covariance matrix	$\Sigma_{\hat{X}} = \mathrm{L}^{\mathrm{T}}\Sigma_Y\mathrm{L}$

Applying the propagation laws…

$$\hat{X} = \underbrace{(\mathrm{A}^T\mathrm{A})^{-1}\mathrm{A}}_{\mathbf{L}^T} \cdot Y$$

= linear estimator

$$\mathbb{E}(\hat{X}) = (\mathrm{A}^T\mathrm{A})^{-1}\mathrm{A}^T\mathrm{A}\mathrm{x} = \mathrm{x}$$

= unbiased estimator

expectation = true value

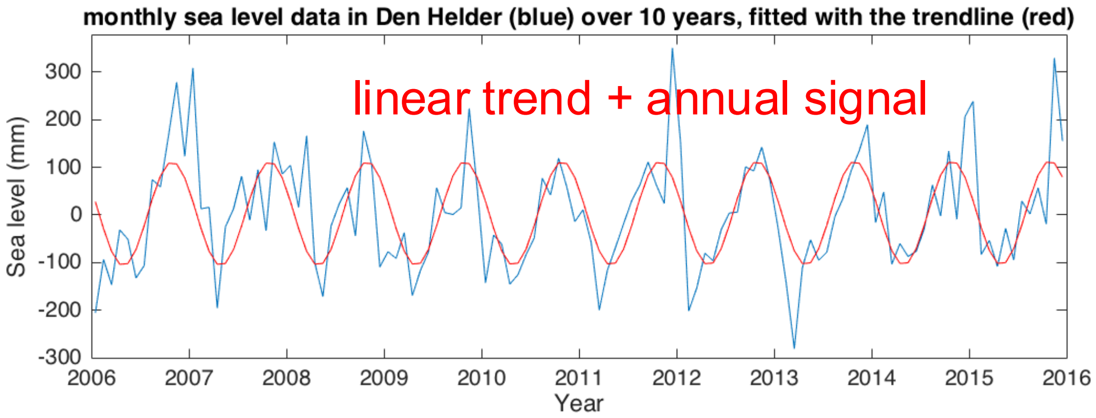
TUDelft

Is least-squares the best way to estimate the parameters (= to fit a model to data)?

Weighted Least-Squares estimation

Modelling, Uncertainty and Data for Engineers

TU Delft



Least-squares…

- Linear model: $$\mathrm{y} = \mathrm{A}\mathrm{x} + \epsilon$$
- Objective $$\min_{\mathrm{x}}(\epsilon^T\epsilon) = \min_{\mathrm{x}}(\mathrm{y} - \mathrm{A}\mathrm{x})^T(\mathrm{y} - \mathrm{A}\mathrm{x})$$
- Solution: $$\hat{\mathrm{x}} = \left(\mathrm{A}^T\mathrm{A}\right)^{-1}\mathrm{A}^T \cdot \mathrm{y}$$
- … treats all observations equally
- But what if observations are collected with different sensors, with different measurement precision?
  - ➢ only use the observations from the best one? NO
  - ➢ give different weights to the observations? YES

TUDelft

Least-squares…

- Linear model: $\mathrm{y} = \mathrm{Ax} + \epsilon$

- Introduce a weight matrix W

- Objective: $\min_{\mathbf{x}}(\epsilon^T W \epsilon)$

- For example with a diagonal weight matrix:

$$\epsilon^T W \epsilon = \begin{bmatrix} \epsilon_1 & \epsilon_2 & \cdots & \epsilon_m \end{bmatrix} \begin{bmatrix} W_{11} & & & O \\ & W_{22} & & \\ & & \ddots & \\ O & & & W_{mm} \end{bmatrix} \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_m \end{bmatrix} = \sum_{i=1}^{m} W_{ii} \cdot \epsilon_i^2$$

TU Delft

An observation with a larger weight is supposed to have a smaller error; this is considered in this minimization problem

Keep the application in mind!

Decisions to be made based on monitoring and sensing:

- can we safely continue with gas extraction / water injection or extraction / CO_2 sequestration?
- do we need to construct higher dikes based on sea level rise predictions / observed deformations?
- do we need to evacuate a region due to risk of a landslide, volcano eruption, tsunami, …?
- is railway maintenance needed?
- is a safe underkeel clearance of ships approaching Rotterdam guaranteed?
- are motions of bridge within safety margins?
- … (etcetera etcetera etcetera)

Need proper data processing and quality assessment of the results

In this part: focus on sensing and monitoring applications

Estimation principles also needed for model verification and validation, regression analysis, machine learning