Data for science
Signal Processing

Practical aspects

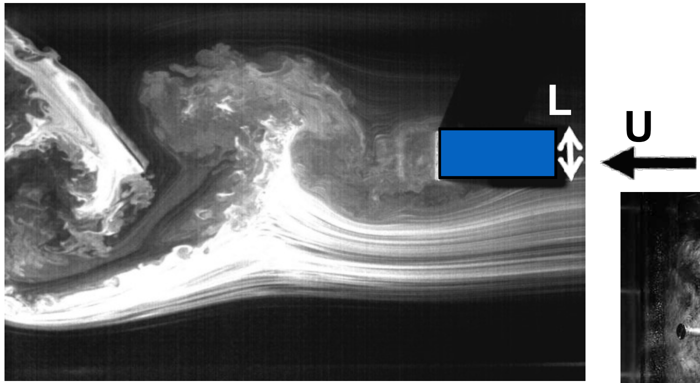
Romain MONCHAUX

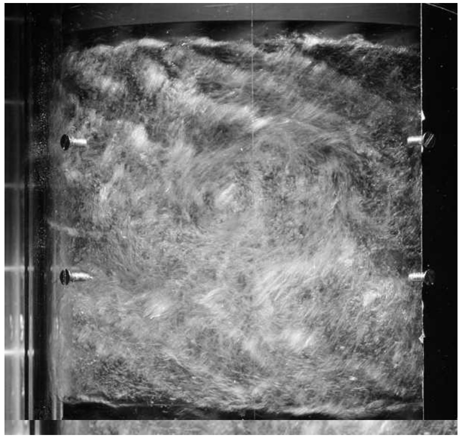
IMSIA, Institute for Mechanical Sciences and Industrial Applications

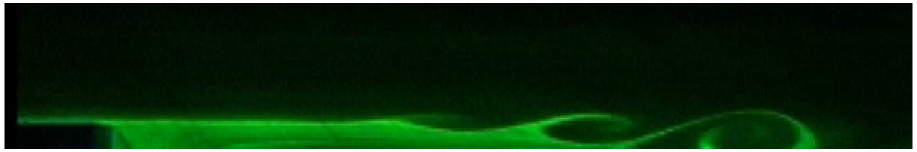
ENSTA – Paris

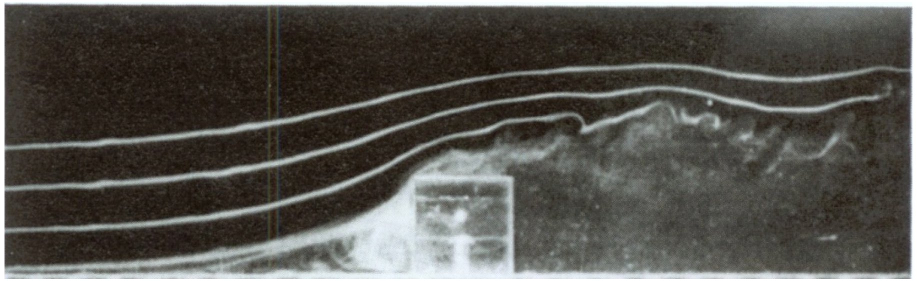
Institut Polytechnique de Paris

Motivations



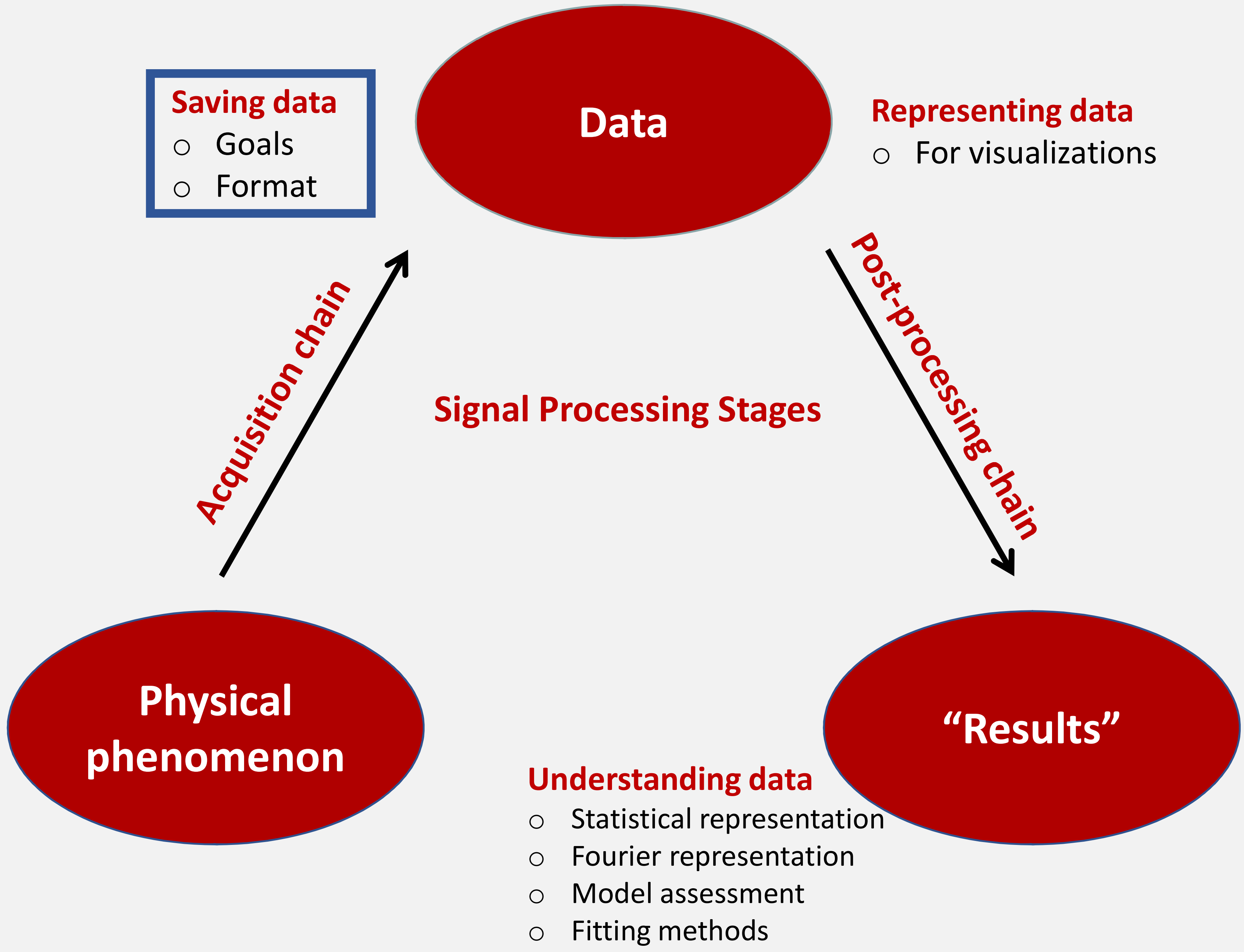




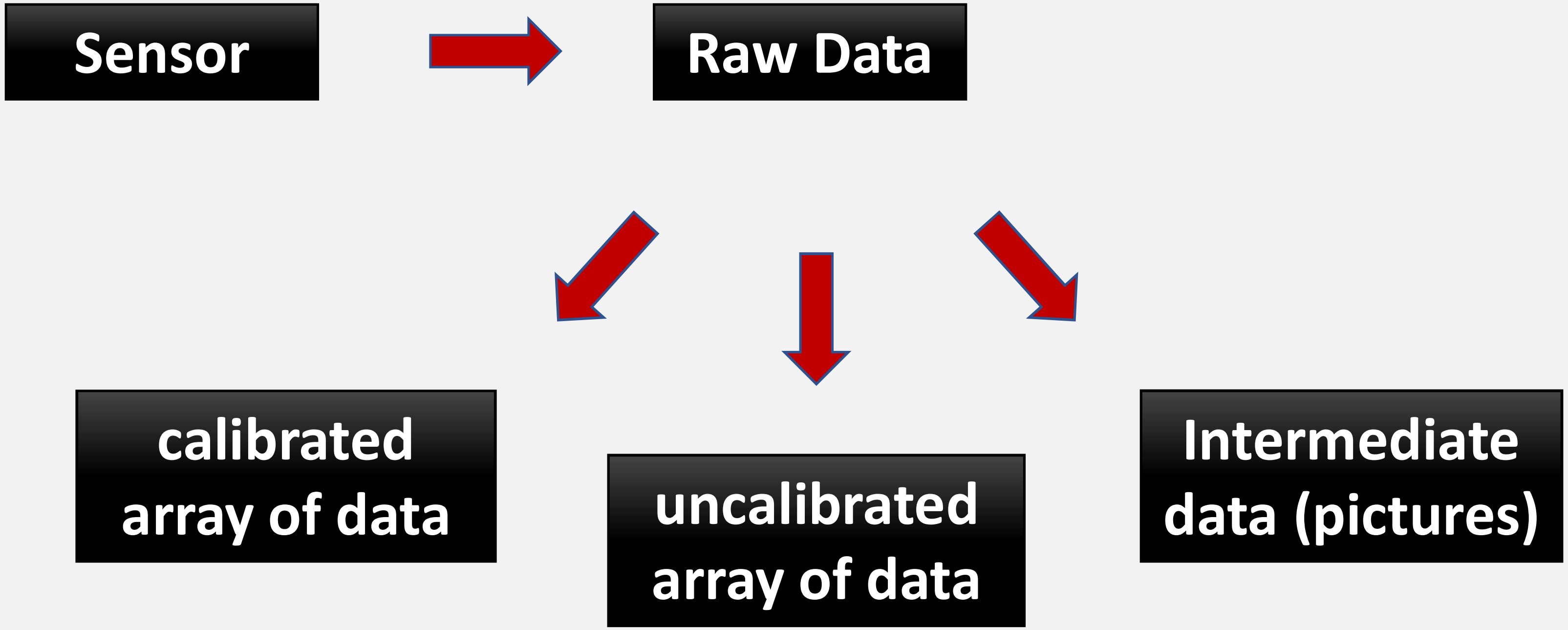


→ **Visualise, quantify, caracterise**

Motivations - Outlines



Data for science: saving data



Wish list

- accessible
- portable
- understandable
- long-life
- space saving
- ...

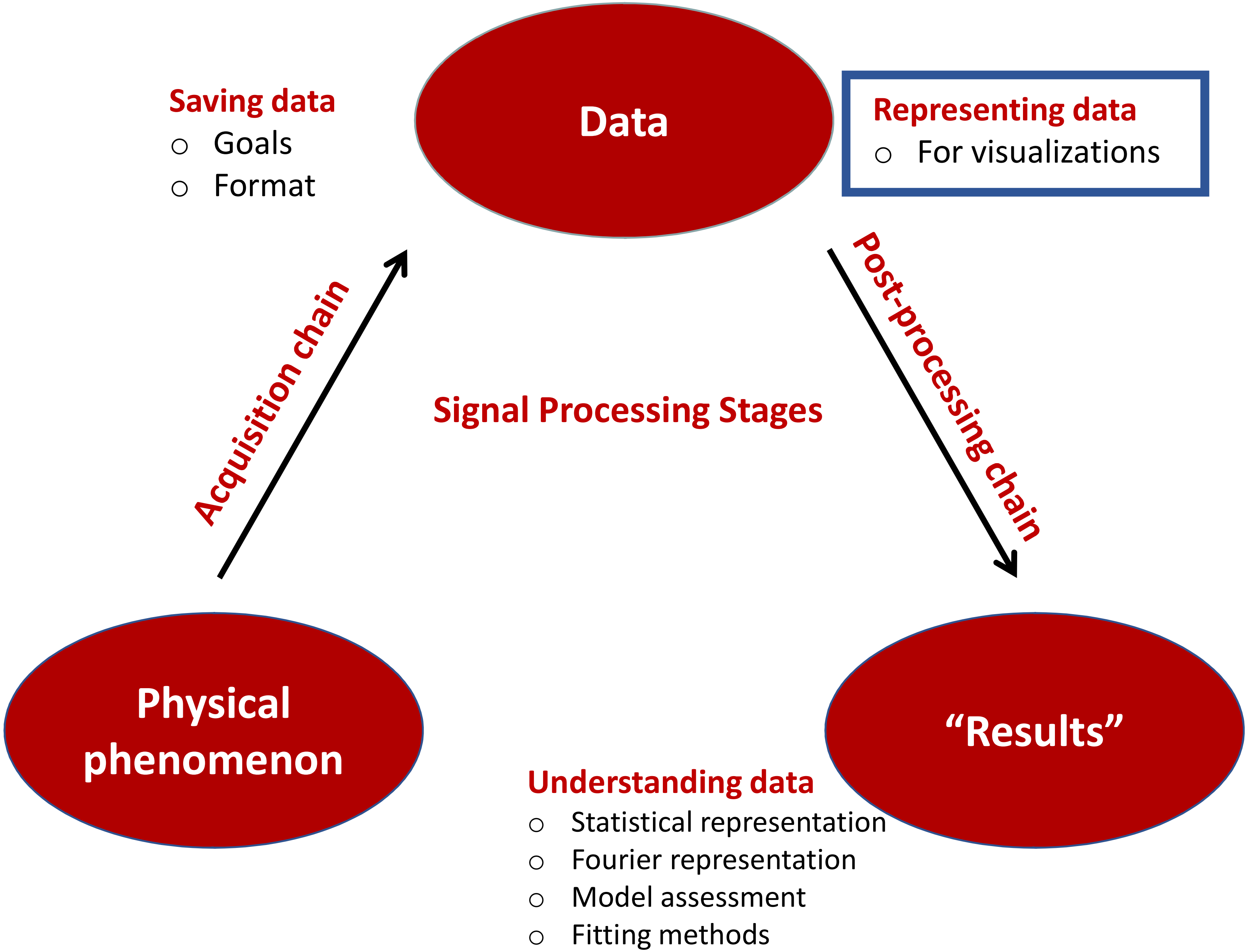
Data representation

- Text files (.csv, .txt, ...)
- Binary files
- Proprietary formats
- .hdf5 or equivalent
- ...

Data location

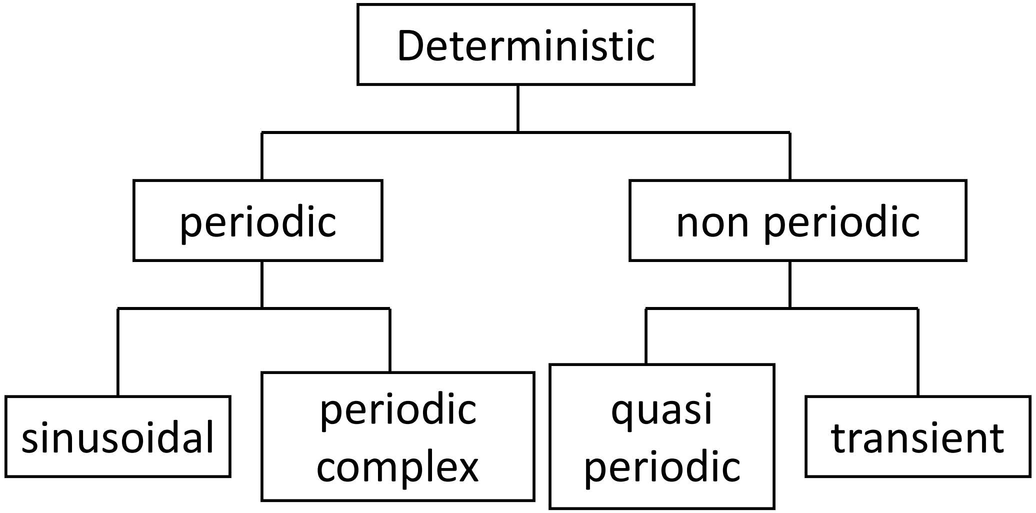
- on computer
- on hard drive
- on servers
- on cloud
- ...

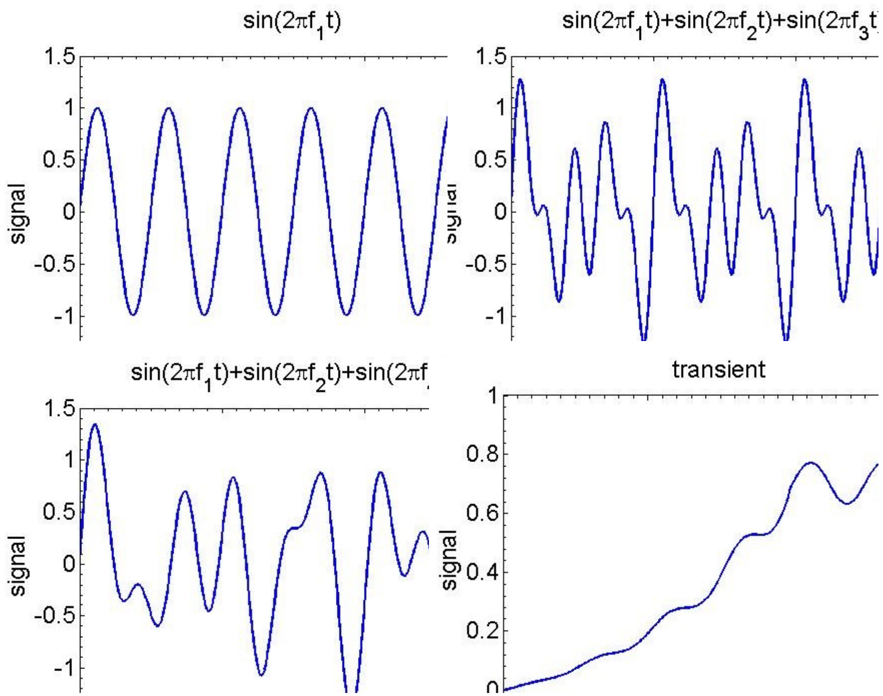
Motivations - Outlines

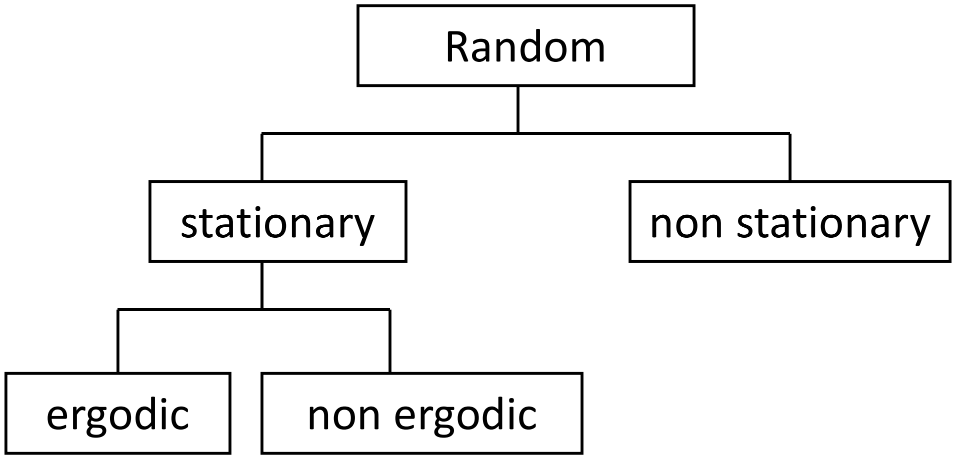


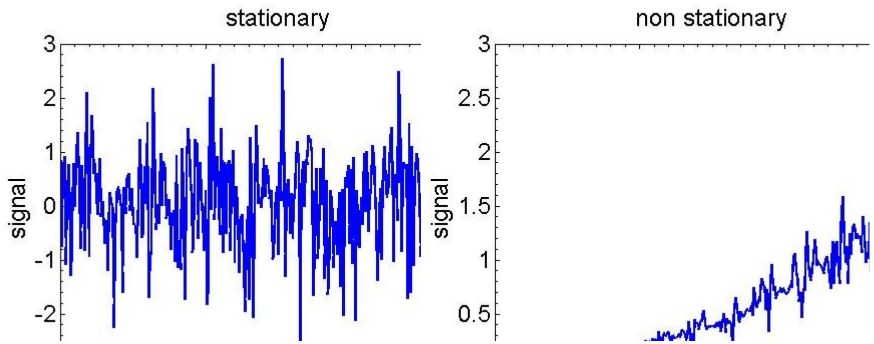
Data for science: representing data

Deterministic vs random signals



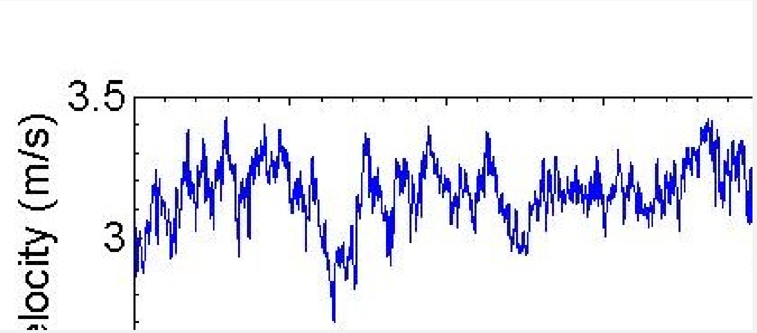




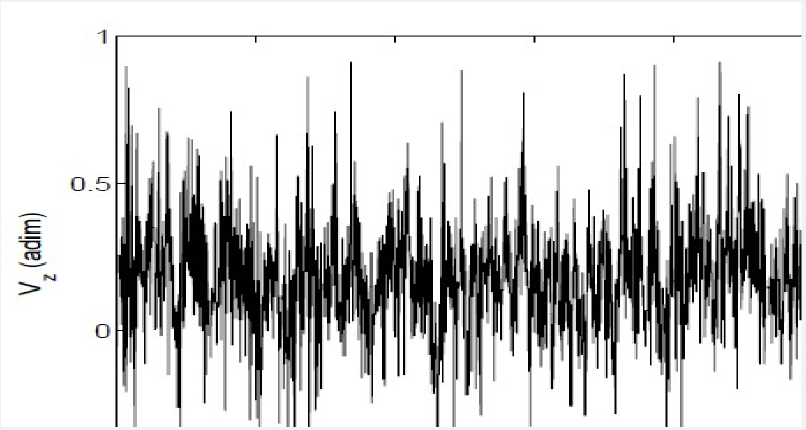


Data for science: visualising data

1D signal: easy

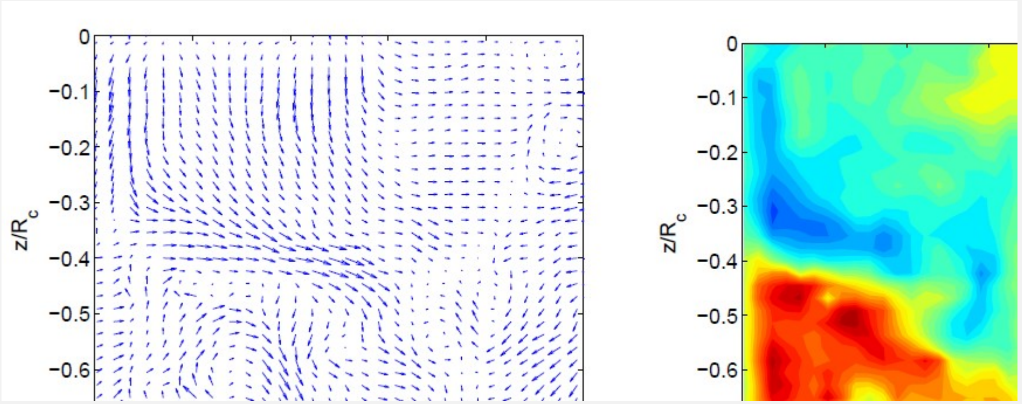


Turbulent jet: hot-wire acquisition

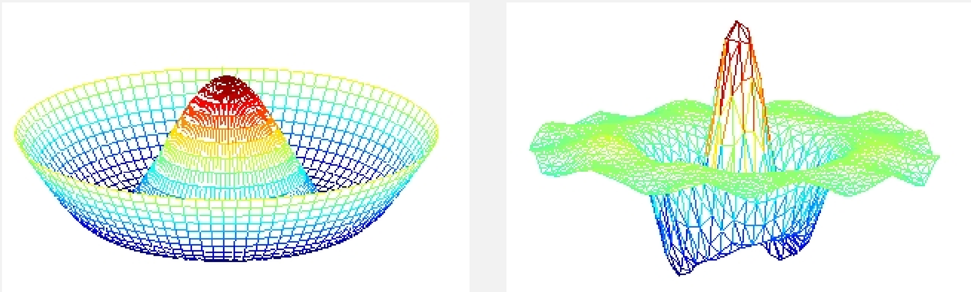


von Karman flow: LDA acquisition

2D signal: more tricky



von Karman flow: stereoscopiv PIV acquisition

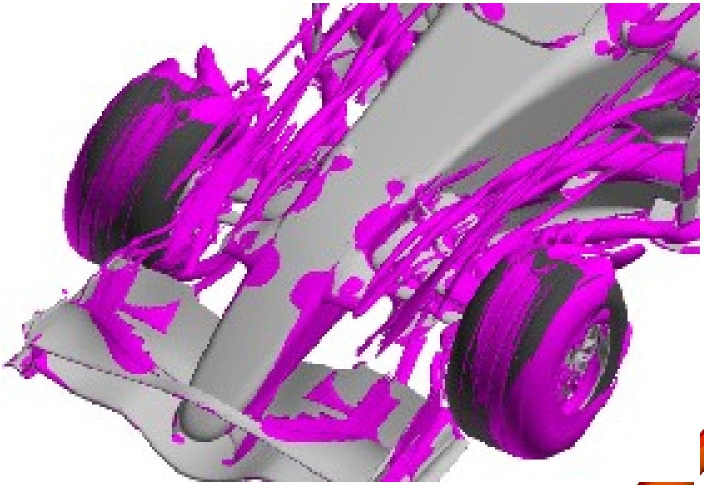


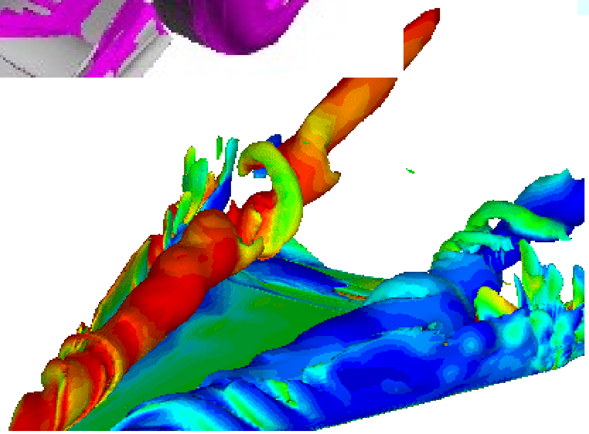
Vibrating flat plates: from Laser vibrometer measurements

What about 3D?
How to represent time variation?

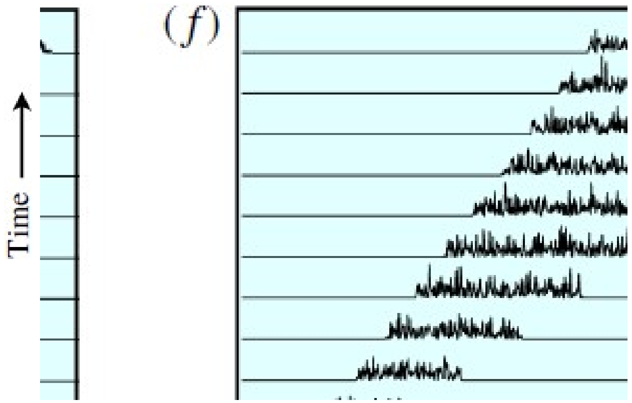
Data for science: visualising data

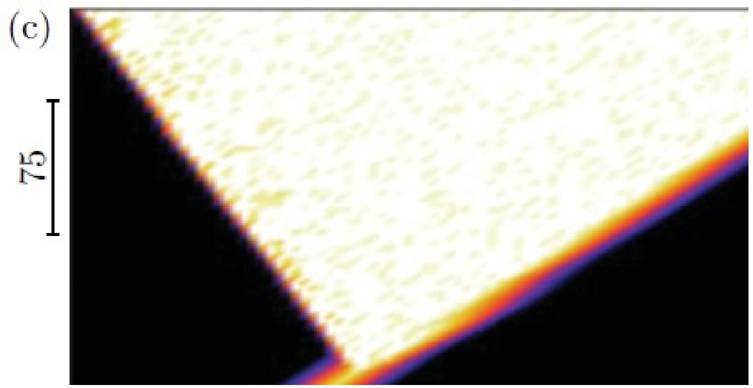
3D data





Time variations of 2D data



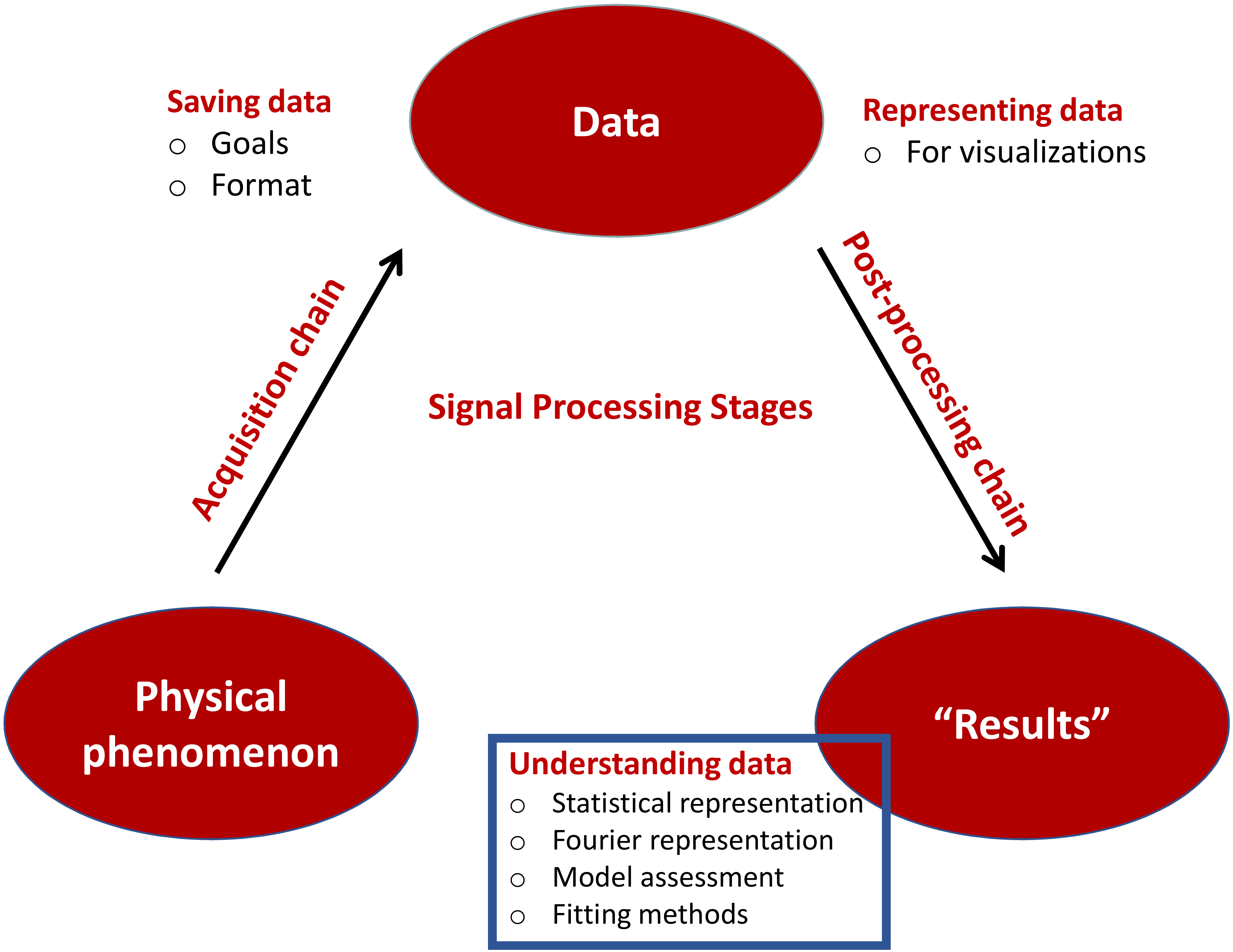


Matplotlib gallery: a good overview

https://matplotlib.org/3.2.1/tutorials/introductory/sample_plots.html

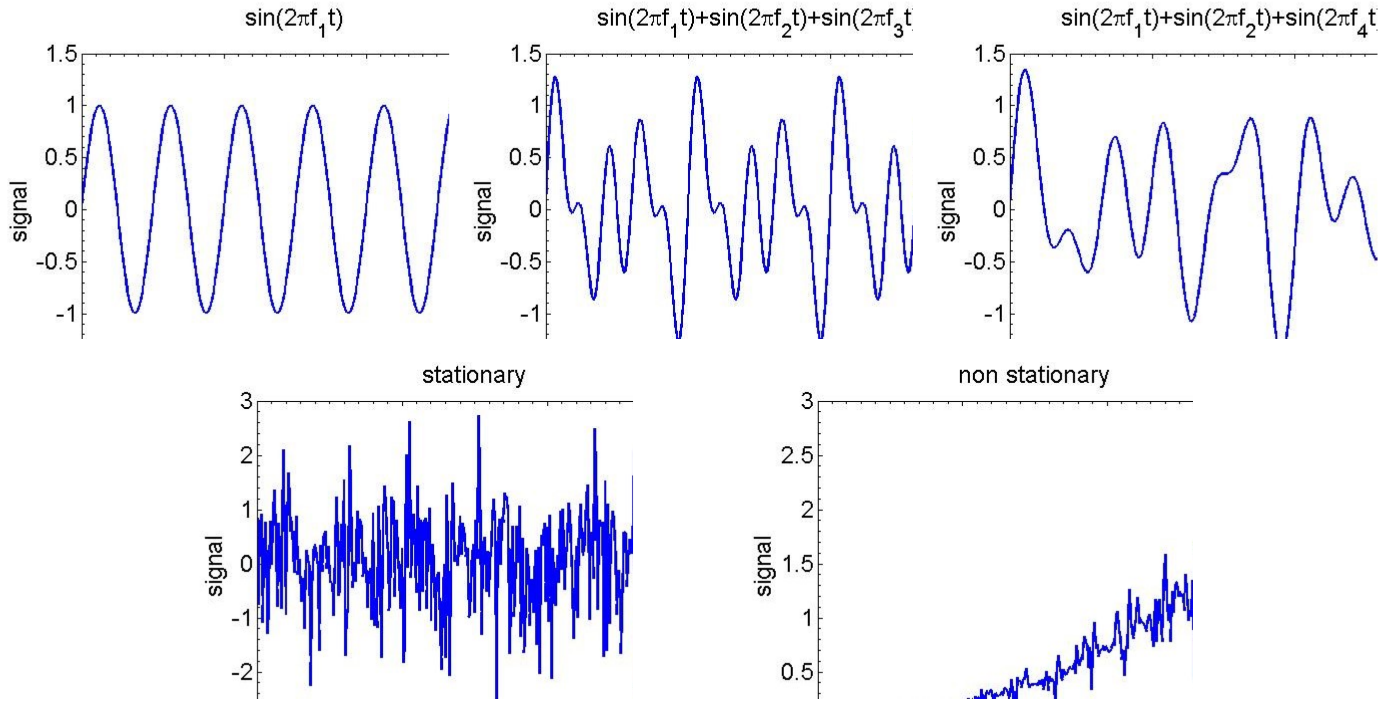
https://matplotlib.org/3.2.0/gallery/index.html

Motivations - Outlines



Data for science: understanding data

Why post-processing?



Statistical quantities

- average, standard deviation
- correlations

Probability density functions:

- fluctuations asimetry
- high order moments
- complex behaviour

Spectral content (Fourier analysis)

- very rich informations

Drift:

- temporal or spatial
- data quality, non trivial long term effects

A historical example

Free fall law

What is the movement of body dropped without initial velocity?
How does their velocity change with time?

Experimental investigation
Galileo experiment

Empirical laws

Test of theoretical prediction
Newton's law
$F = m\, dv/dt$

Theory invalidation
… or not

Data for science: understanding data

Copernician revolution

Before	After
Experimental investigation	**Test of theoretical prediction**
- Finding empirical laws	- First principles
- Feeding theoretical developments	- Development of theoretical models
- Revealing unknown behaviors	- Experimental tests

and when no theory is available

or when all existing theories fail

Data for science: understanding data

Statistical quantities

- average, standard deviation
- correlations

Work on Jupyter:

- Load a simple temporal serie
- Build an estimate for the average and the standard deviation
- Compare to Python built-in function
- Estimate convergence speed to the « real value »

Data for science: understanding data

Statistical quantities

- average
- standard deviation

Non biased estimators:

Average:

$$\overline{p} = \frac{1}{M} \sum_{n=0}^{M-1} p[n]$$

Standard deviation:

$$p_{rms} = \Bigg(\frac{1}{M-1} \sum^{M-1} (p$$

Convergence

○ faster with uncorellated data => obvious
○ as square root of number of data points
○ be carefull: work with uncorrelated samples or take more points!

Data for science: understanding data

Probability density functions:

- fluctuations asimetry
- high order moments
- complex behaviour

Building an histogram, a PDF:

-Nb Bin choice

-Range of bins choice

Data for science: understanding data

Probability density functions:

- fluctuations asimetry
- high order moments
- complex behaviour

Important parameters:

- ○ number of bins not too high
- ○ number of bins not too low
- ○ again: take into account correlations

Data for science: fitting model to data

Curve fitting

- interpolating data
- smoothing data
- testing known laws (test of predictions)
- empirical exploration
- linear vs non linear fitting

Data for science: fitting model to data

General principles

- experimental data points: y_i^{exp}
- calculated values: y_i^{cal}
- minimizing distance between data points and calculated function
    - least-square calculation vertically
    - least-square calculation perpendicularly to the curve
- Reliability factor (R)

$$\mathrm{R_{wp}} = \sqrt{\frac{\sum_i w_i \cdot (y_i^{\mathrm{exp}} - y_i^{\mathrm{cal}})^2}{\sum_i w_i \cdot y_i^{\mathrm{exp}\,2}}}$$

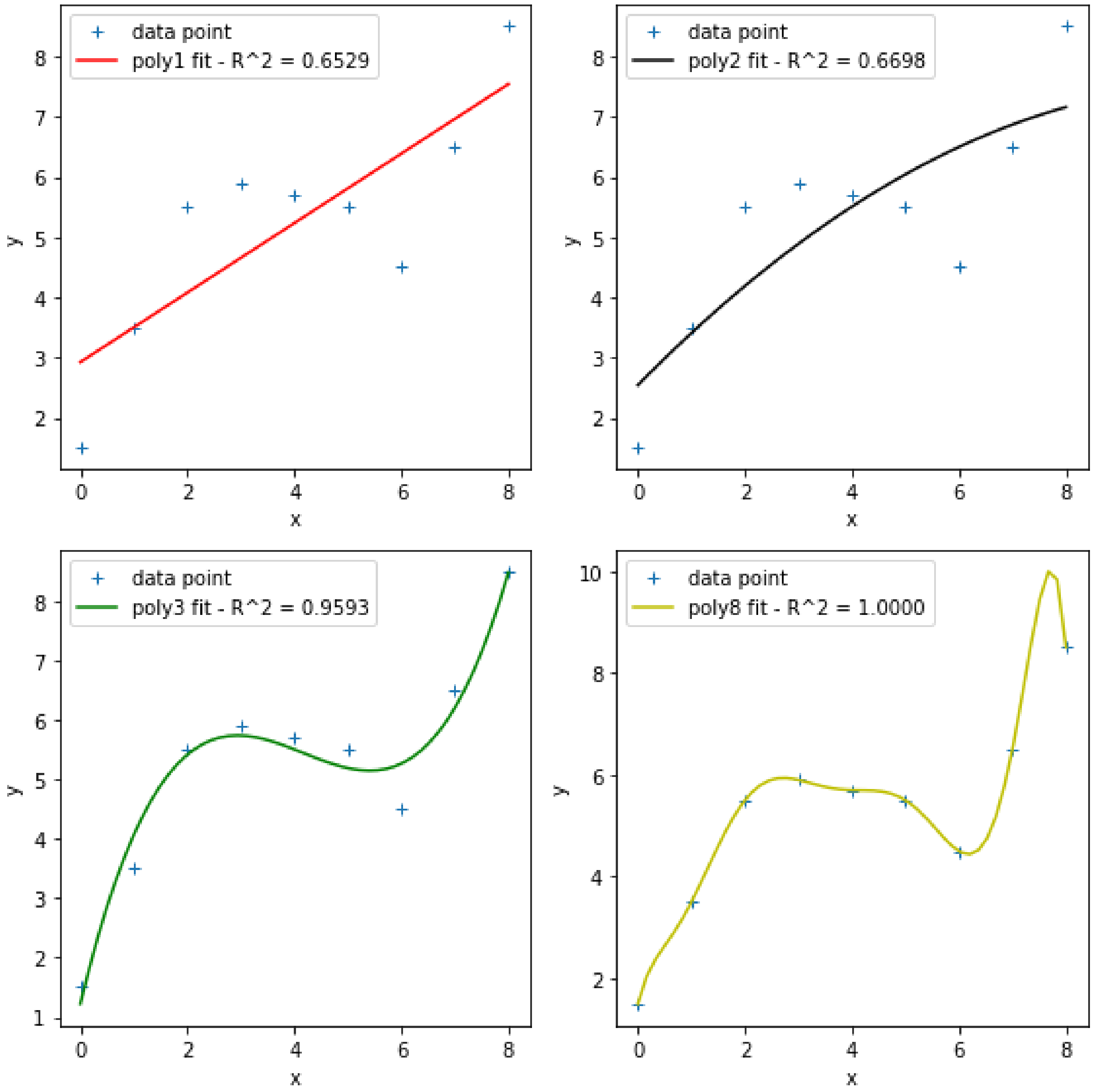
weighted

$$\mathrm{R} = \sqrt{\frac{\sum_i \cdot (y_i^{\mathrm{exp}} - y_i^{\mathrm{cal}})^2}{\sum_i \cdot y_i^{\mathrm{exp}\,2}}}$$

or not weighted

More parameters: danger
Polynomial example

Data for science: fitting model to data



Too many parameters:

Danger