Basics of Signal Analysis:

Measurement, Sampling, Pre-processing

Olaf Hauk

MRC Cognition and Brain Sciences Unit
olaf.hauk@mrc-cbu.cam.ac.uk

Data Sampling in Space and Time

Example: EEG data

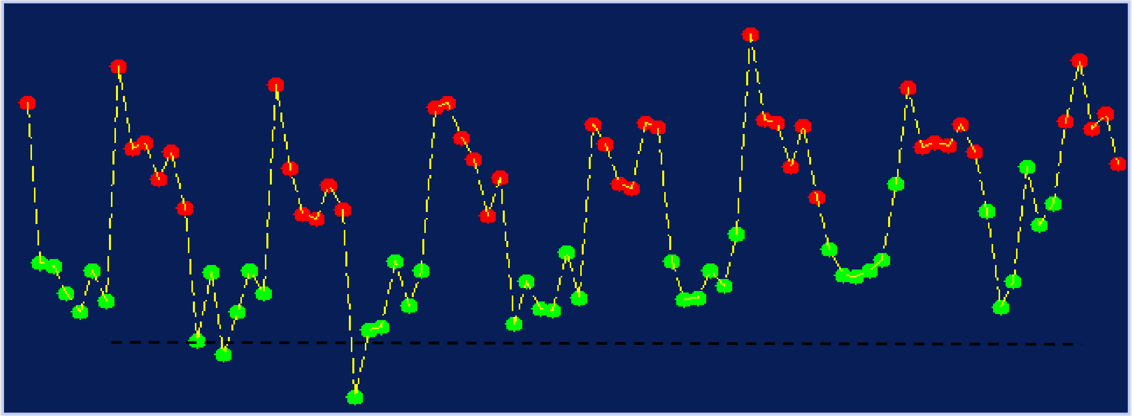
Electrodes
Rows

Time Points
Columns

Data Sampling in Time

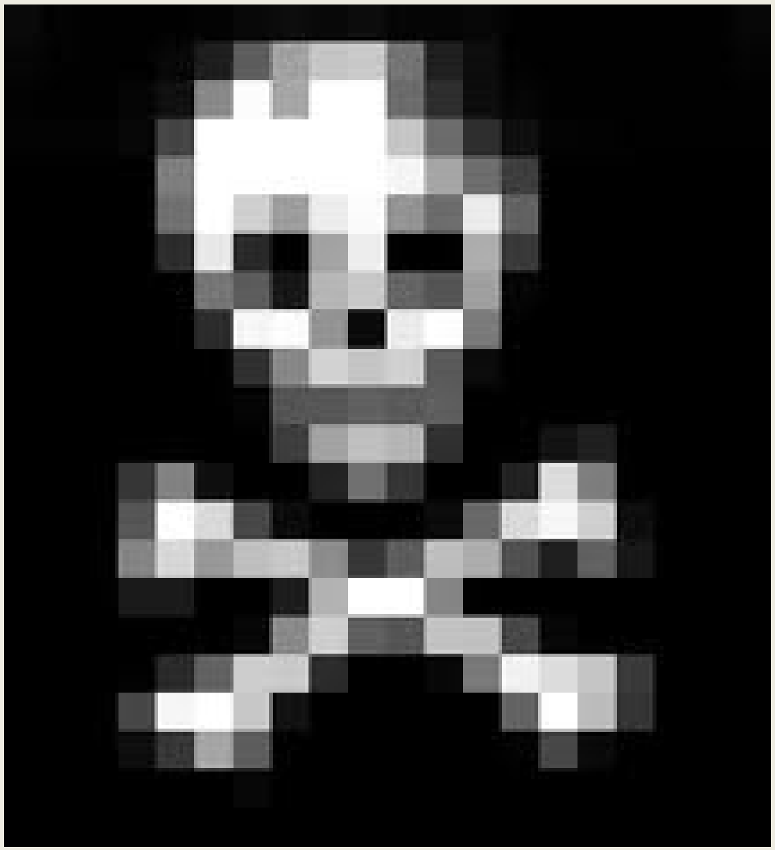
Example: fMRI data

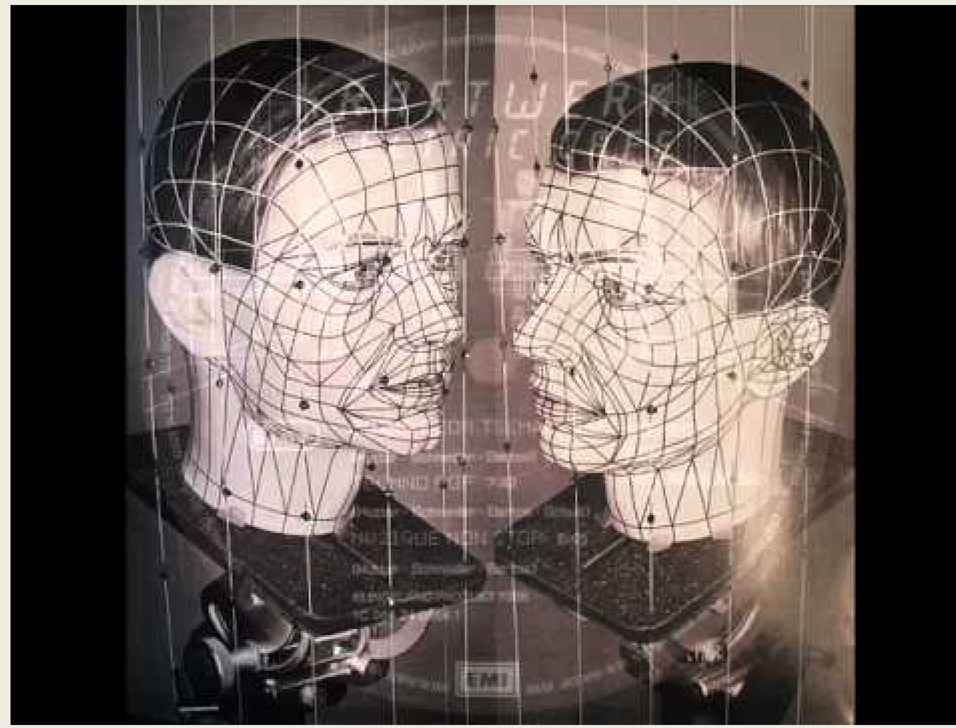
voxel timeseries



Data Sampling in Space

Example: Pictures





What is Data Sampling?

- We are often interested in a continuous signal that we would like to capture and analyse (brain activation in time and space, 2D image, 3D object etc.)

- Digital signal processing can only handle discrete numbers, i.e. vectors, matrices etc.

- Therefore, we need to represent the continuous signal by means of representative samples. This is sometimes called “discretisation”.

- Sampling should provide the information necessary for the intended analysis, while at the same time allow for efficient processing

Basic Concepts

- It is usually most convenient to sample **equidistantly**, i.e. neighbouring samples have the same distance to each no matter at what point of the sample they are

- **Sampling Rate/Frequency**: How densely do we take samples? For example:
  100 samples per second -> 100 samples/s -> 100 Hz
  10 samples per centimetre -> 10 samples/cm
  100 samples (“pixels”) per square centimetre -> 100 samples/cm^2

- **Sampling Interval/Distance**: How far apart are the samples (in time, space etc.)?
  100 Hz -> (1/100)*1s = 0.01 s = 10 ms
  10 samples/cm -> (1/10)*1 cm = 0.1 cm = 1mm
  100 samples/cm^2 = (1/100)*1 cm^2 = 0.01 cm^2 = (0.1*0.1) cm^2 = 1 mm^2

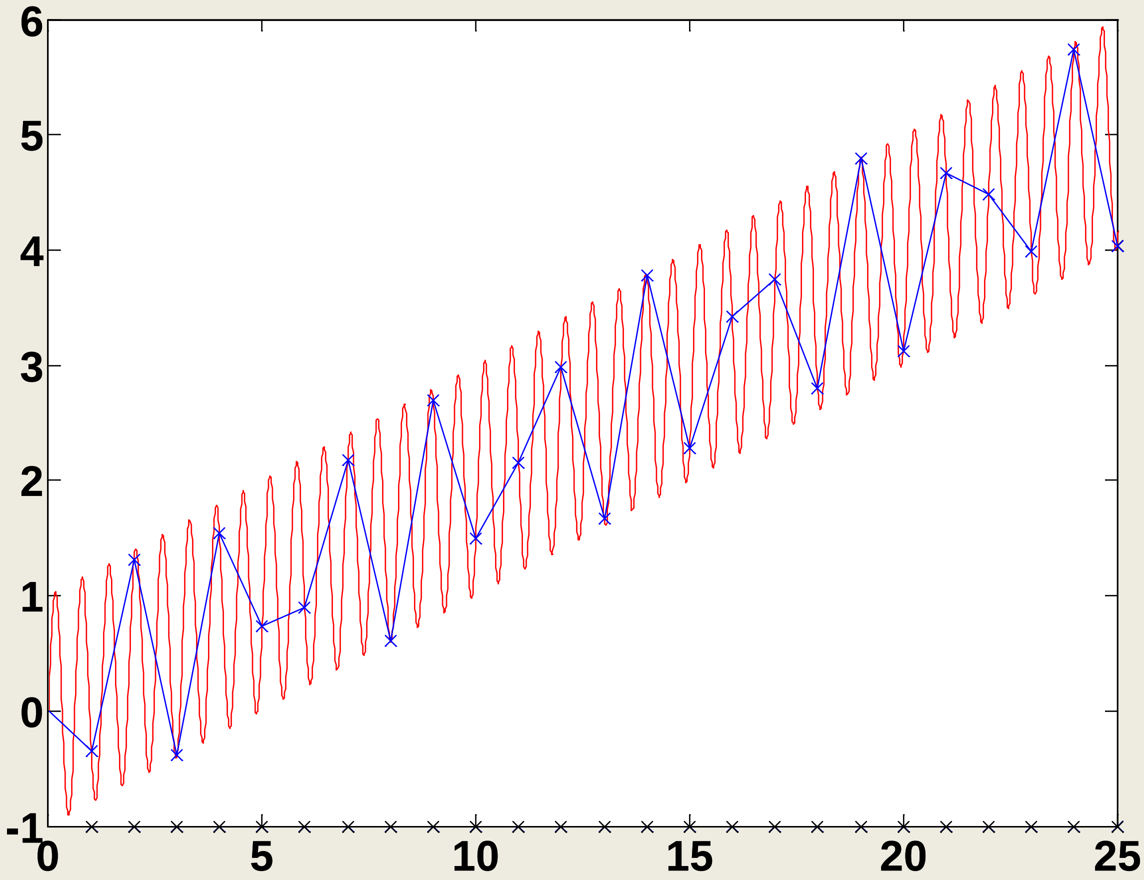
- **Sampling depth (quantisation)**: For one particular sample, how many different values can we separate in digital representation?
  - “2 bit”: Either 1 or 0, we can only separate 2 values (e.g. Black/White)
  - “8 bit”: 1 2 4 8 16 32 64 128 => 256 different values
    [1/0 1/0 1/0 1/0 1/0 1/0 1/0 1/0]

- **Sampling range:** What are the maximum/minimum values we can sample?

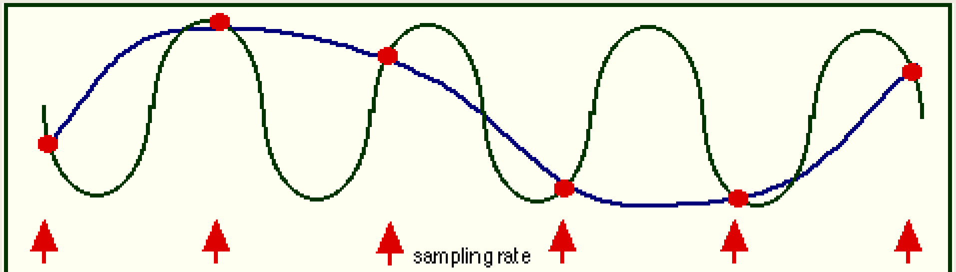
- **Resolution/precision:** Range divided by depth
  For example: Range +/- 10 μV, 8 bit sampling depth => 20/256 ≈ 0.08 μV

Example

Downsampling Can Lead to "Aliasing"



Downsampling Can Lead to “Aliasing”



Aliasing produces waveform (blue) at 1/4 the original signal's frequency (green).

Nyquist(-Shannon) Sampling Theorem

If you sample a signal with a sampling rate of *X* Hz, make sure it does not contain frequencies above *X/2* Hz.

X/2 is the "Nyquist Frequency" (half of the sampling frequency).

The largest frequency in your signal should be smaller than the Nyquist Frequency.

Good software takes care of this during acquisition or analysis, by applying appropriate filters.
Often data are filtered with a cut-off frequency well below the Nyquist frequency (e.g. X/3), to account for other sources of inaccuracies.

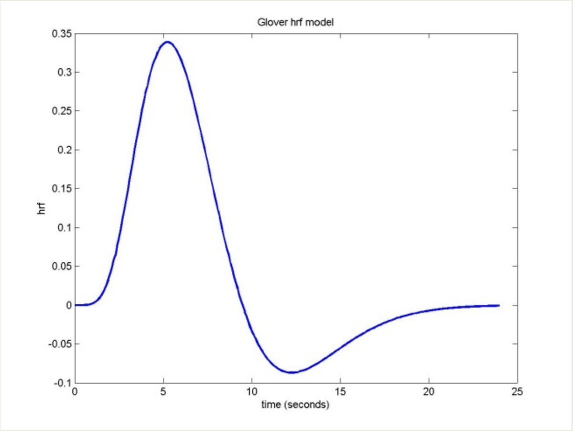
But note:
If you sample with frequency *X*, you cannot get information about signals with frequencies above *X/2*.

Examples for Sampling Theorem

fMRI:

We typically sample every 2 seconds
(0.5 Hz with TR=2s).

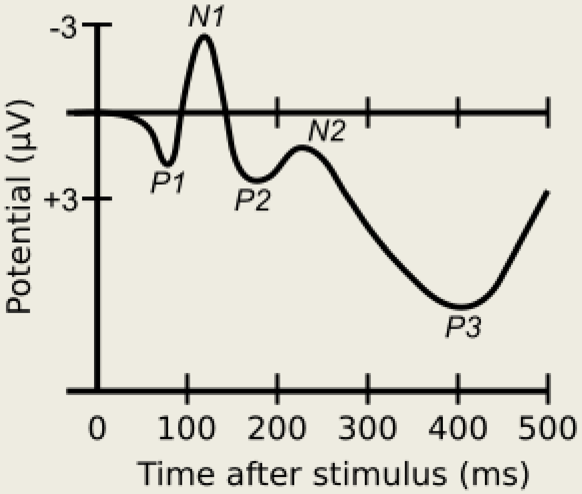
=> Nyquist Frequency 0.25 Hz



EEG/MEG:

A typical sampling rate is 500 Hz
(sampling distance 2 ms)

=> Nyquist Frequency 250 Hz



Note: Artefacts should be sufficiently sampled too, in order to deal with them later
(e.g. important for combined EEG+fMRI)

Missing Data: Interpolation

Missing data points can only be "corrected" if we have some model for our data (e.g. about its smoothness)

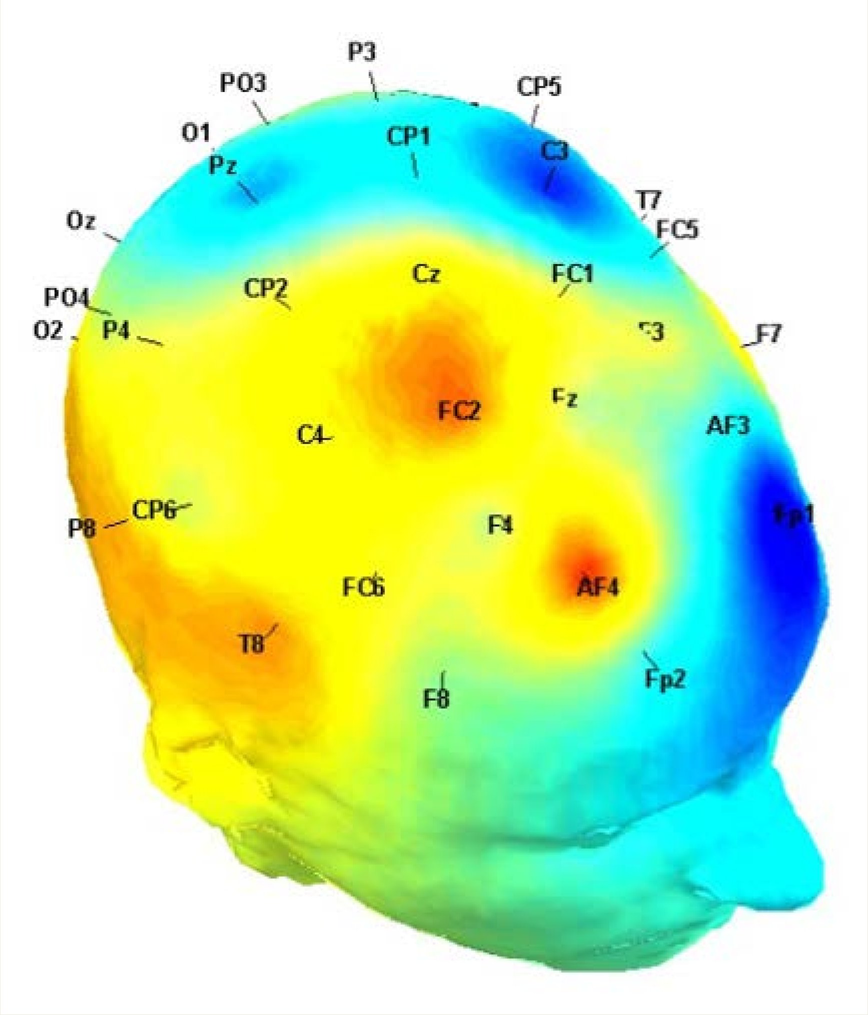
One possibility is: Leave missing data points out of your analysis (often not practical)

Interpolation: Replace missing data points by the best guess

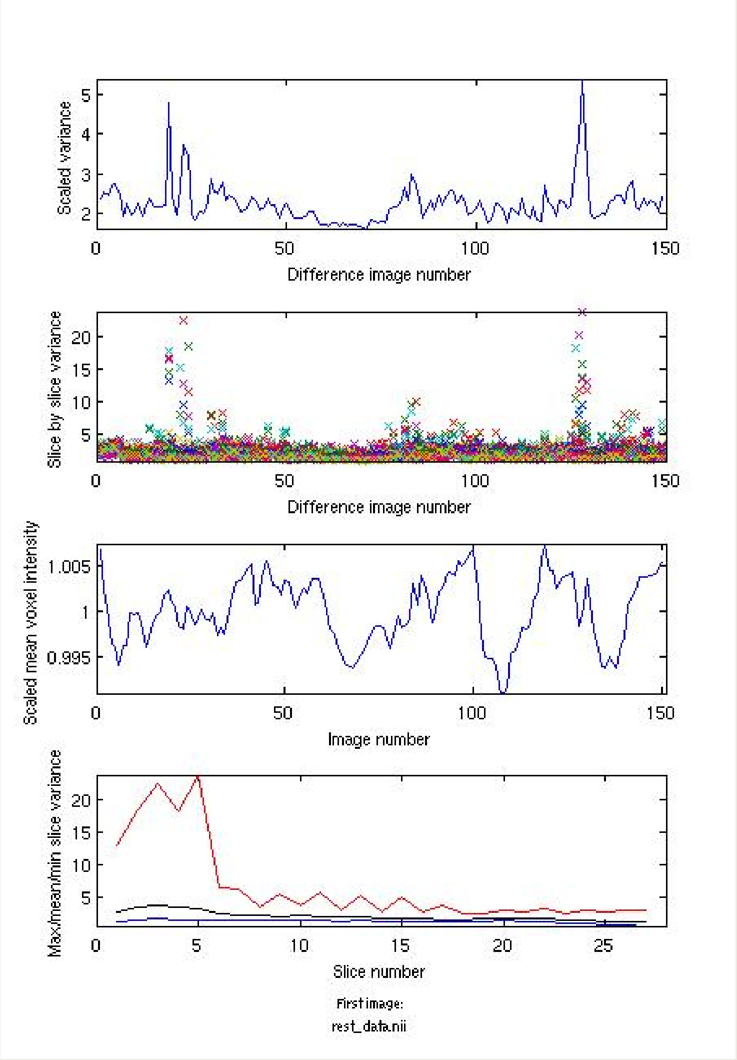
Note:
Interpolation does not add information to your data.

Missing Data: Interpolation

Faulty electrodes



Faulty scans



Example

Data Quality: Signal-to-Noise Ratio (SNR)

Signal-to-Noise ratio:
Compare the level of your "signal" to the level of your "noise"
(define "signal" and "noise" first)

Common definition for SNR:
Divide power (variance) of signal by power (variance) of noise

Other definitions possible:
Divide amplitude of signal by standard deviation of noise
Divide root-mean-square (RMS) of signal by RMS of noise

$$SNR = \frac{P_{Signal}}{P_{Noise}}$$

Decibels: $$SNR_{dB} = 10\log_{10}\frac{P_{Signal}}{P_{Noise}} = P_{Signal,dB} - P_{Noise,dB}$$

Example

Noise and Error Propagation

Noise can take on many forms, for example:

- Inaccuracies of measurement equipment
- Interference from artefact sources
- Modelling errors

Any transformation of your data will be affected by noise, and may amplify it

For example: Subtracting/Adding data sets with equal variance doubles the variance

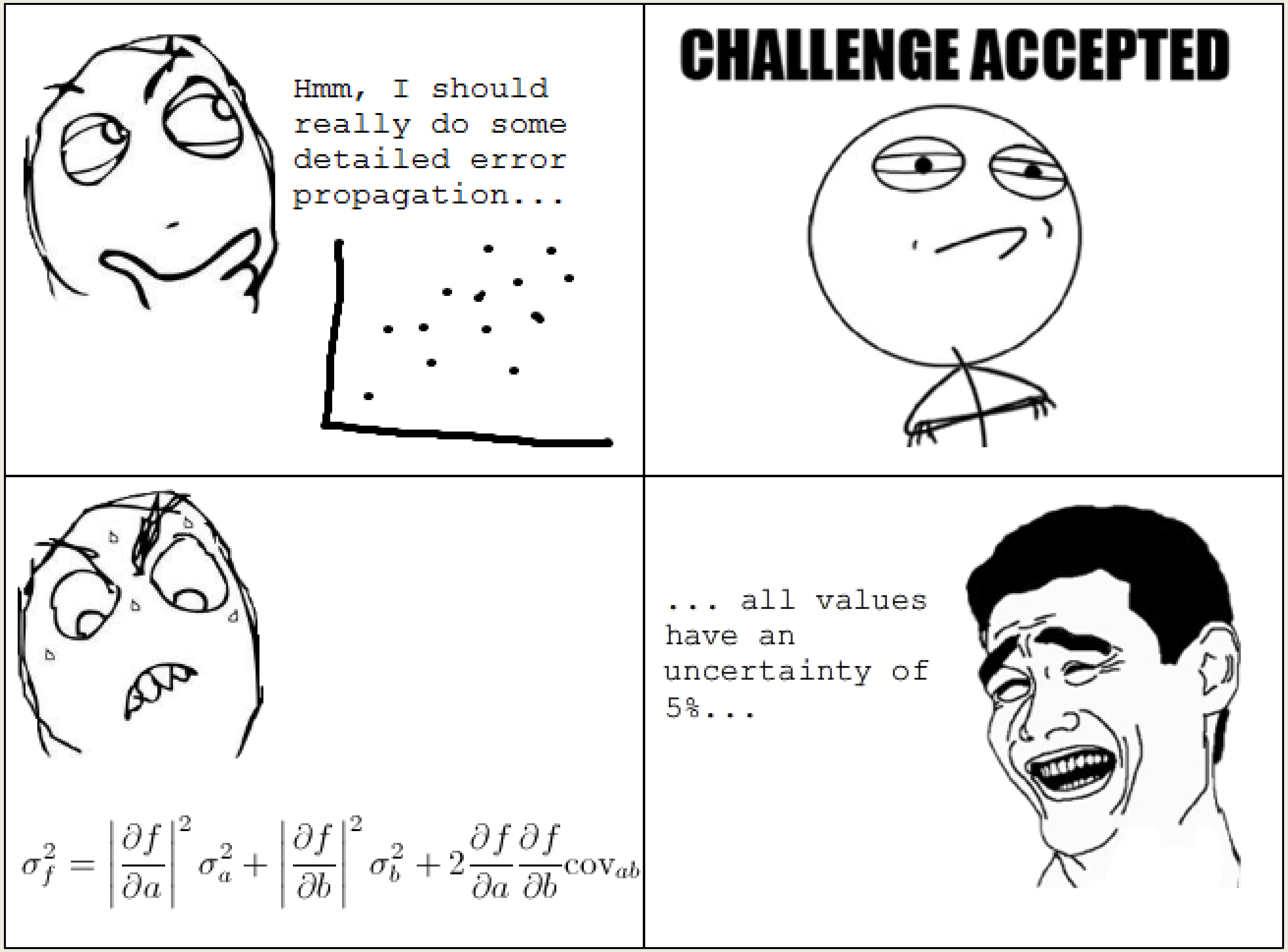
If the operation is more complex, the effect of noise will probably be more complex, possibly with disastrous consequences (“error propagation”):

Noise and Error Propagation

Example:

- You can only measure x, but you want y=1/x
- Noise in x fluctuates around 0 (let’s say between -0.5 and 0.5)
- If you measure x=10, then y can be between 1/(9.5)=0.105 and 1/(10.5)=0.095
- If you measure x=0.5, then y can be between 1/0=Inf and 1/1=1

Noise and Error Propagation



http://i.imgur.com/IAude.png

Example

Data Smoothing

Because of noise and error propagation, it’s usually a good idea to remove noise and artefacts as early as possible during data processing – but don’t throw out the child with the bathwater

First processing steps are usually smoothing or filtering

The ideal filtering depends on properties of your data – which you may know or not

Filtering may produce awkward effects at the edges of your data – visualisation is important

Example