Statistical Analysis of the Experimental Data

It's not a Pain, but a Principle!

Why are we doing this?

Engineering Statistical Analysis:

• Engineers need statistics to establish the performance of equipment. <u>We can't be wrong</u> (Safety, performance..)

• News, economists, and politicians are heavy with statistics, primarily predicting the future based on past history or polls. <u>Generally, wrong.</u>

VOCABULARY: “Forecast” and “Prediction”

Prediction: A definite, specific statement about when and where an event will occur.

(Most economists, sports writers, and politicians make Predictions)

Forecast: A probabilistic statement related to a period of time. Weather forecasts are the most familiar.

Engineering: The probability of this bearing failing after 10,000 hours service is 20%.

How do you know that? Experimental data

Data always has uncertainty at some level

Here is a famous example:
Is the universe expanding?

Einstein said "Nonsense" (at first)

The Velocity-Distance Relation among Extra-Galactic Nebulae from Edwin Hubble's original paper published in Proceedings of the National Academy of Sciences on 15 March 1929.

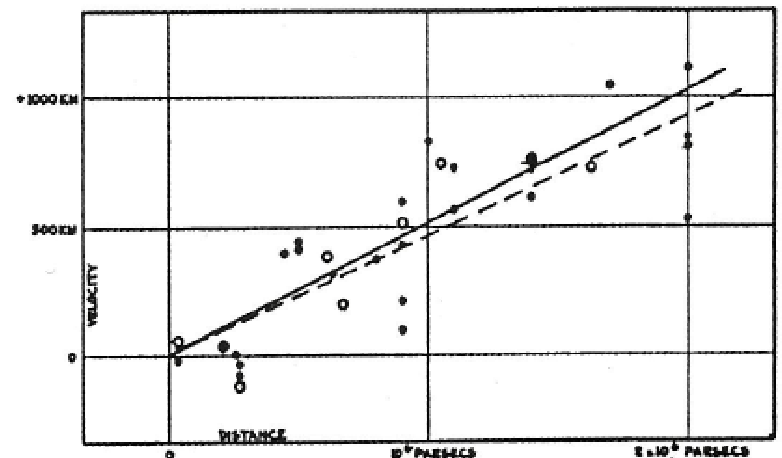


FIGURE 1

Mon. Not. R. Astron. Soc. **000**, 1–6 (2014) Printed 27 May 2014 (MN LaTeX style file v2.2)

Two planets around Kapteyn's star : a cold and a temperate super-Earth orbiting the nearest halo red-dwarf

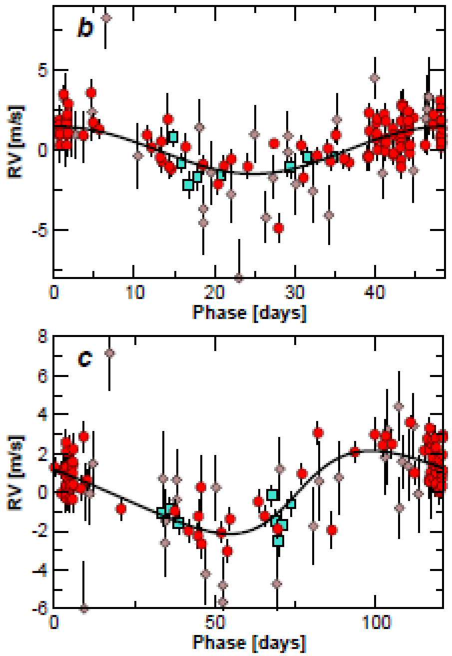


Figure 2. Phase folded Doppler curves to the reported signals with the other signal removed (HARPS are red circles, HIRES are brown diamonds and PFS are blue squares). The maximum likelihood solution is depicted as a black line.

CLOSER TO HOME STATISTICS ARE EVERYWHERE

- Election polls and predictions
- Sports writers and analysts
- Economics – stocks, employment, sales…
- Freshman choices for their major

SAMPLING BIAS (e.g., election outcome)

- How many to ask? (1000? Every voter?)
- Respondents tend to be passionate (not a true random sample)
- Not every respondent ends up voting
- Respondents change their mind...

STATISITICS CAN BE NON-INTUITIVE AND SURPRISING !

WHAT IS THE PROBABILITY THAT TWO OF YOU HAVE THE SAME BIRTHDAY (Month and Day)?

$$P = 1 – (365/366)*(364/366)*(363/366)*\ldots(263/366)$$

(number of terms = number in group)

5	.03
10	.12
20	.41
25	.57
50	.97
99	.99

Error or Uncertainty?

Two components: Biases
Random factors

Biases: Inherent in the instrument and can be calibrated out.
Basic user error: e.g., Recording ^{o}C as ^{o}F, Mpa as KPa

Random: Many sources, varying with time, space and operator
and the instrument itself

Biases can usually be detected and corrected.

Random factors are treated with statistical methods.

Measurement Instruments – where the data comes from

In Fluids and Heat Transfer, primary variables are:

- Temperature & Pressure
- Flow rates
- Power
- Fluid and material properties

Applications:

- Monitor Processes and Operations
- Control Processes and Operations
- Experimental Engineering Analysis

Engineering Analysis - Theoretical vs Experimental

Theoretical:

- Results are of general use based on principles
- Invariably require simplifying assumptions

Experimental:

- Results apply only to the specific system tested
- No simplifying assumptions necessary
- Accurate measurements necessary for true picture

Resolution of conflict?

Need both if at all possible.

Nobody trusts a computer simulation except the guy who did it,

and everybody trusts the experimental data except the guy who did it.

From “Hot Air Rises and Heat Sinks” by
Tony Kordyban

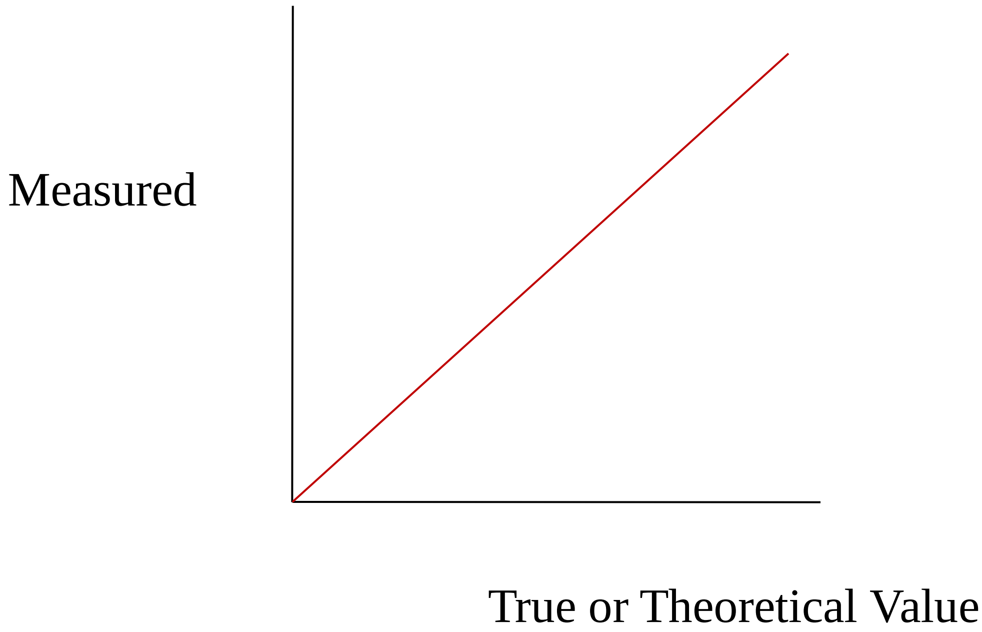
Experimental Data Analysis:

Average Difference?

Root Mean Square?

Standard Deviation?

Propagation of Error?

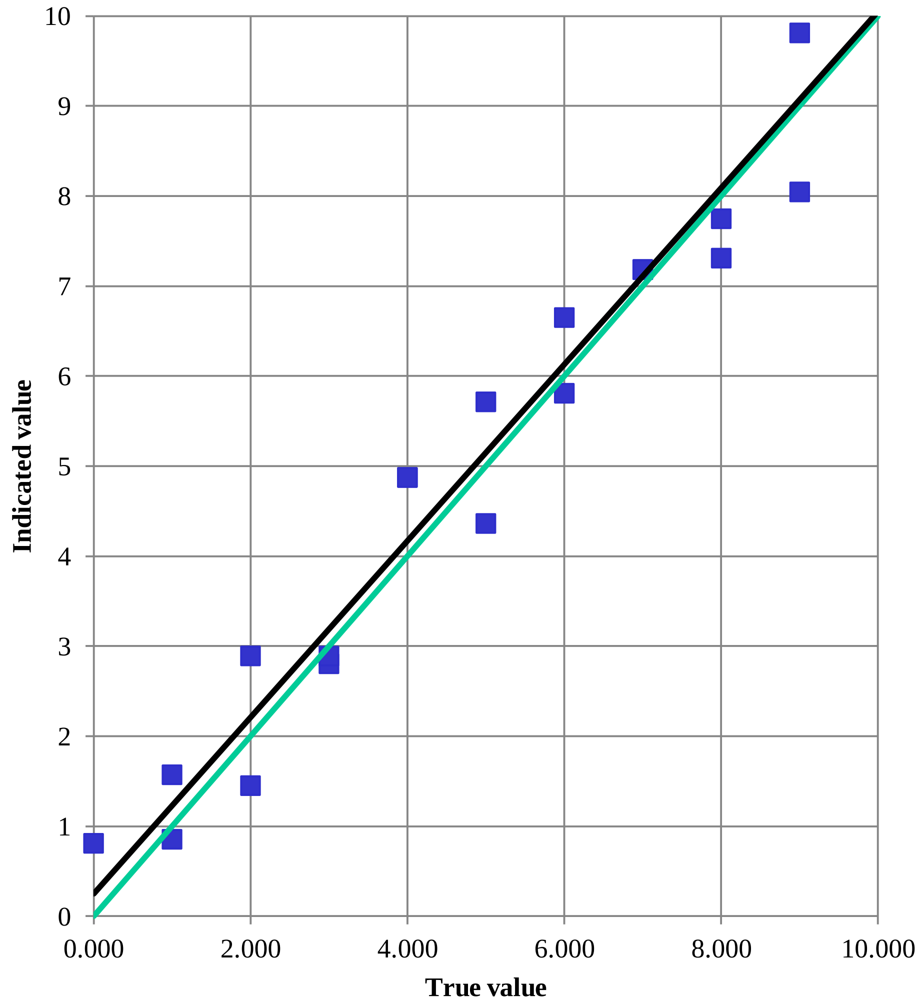


Surely your data points will fall right on the 45^{o} line!

If they do not, your uncertainty analysis should point to possible reasons why. (Bias can often be spotted).

Data from a gage calibration

True Press	45 Degree	Indicated
0.000	0.000	0.809551
0.000	0.000	-0.35478
1.000	1.000	1.571265
1.000	1.000	0.854223
2.000	2.000	2.891452
2.000	2.000	1.450355
3.000	3.000	2.804081
3.000	3.000	2.889775
4.000	4.000	4.874644
4.000	4.000	4.872998
5.000	5.000	5.713672
5.000	5.000	4.362477
6.000	6.000	5.809227
6.000	6.000	6.650267
7.000	7.000	7.184883
7.000	7.000	7.181706
8.000	8.000	7.747654
8.000	8.000	7.310156
9.000	9.000	8.045593
9.000	9.000	9.812118
10.000	10.000	10.09085
10.000	10.000	10.67398

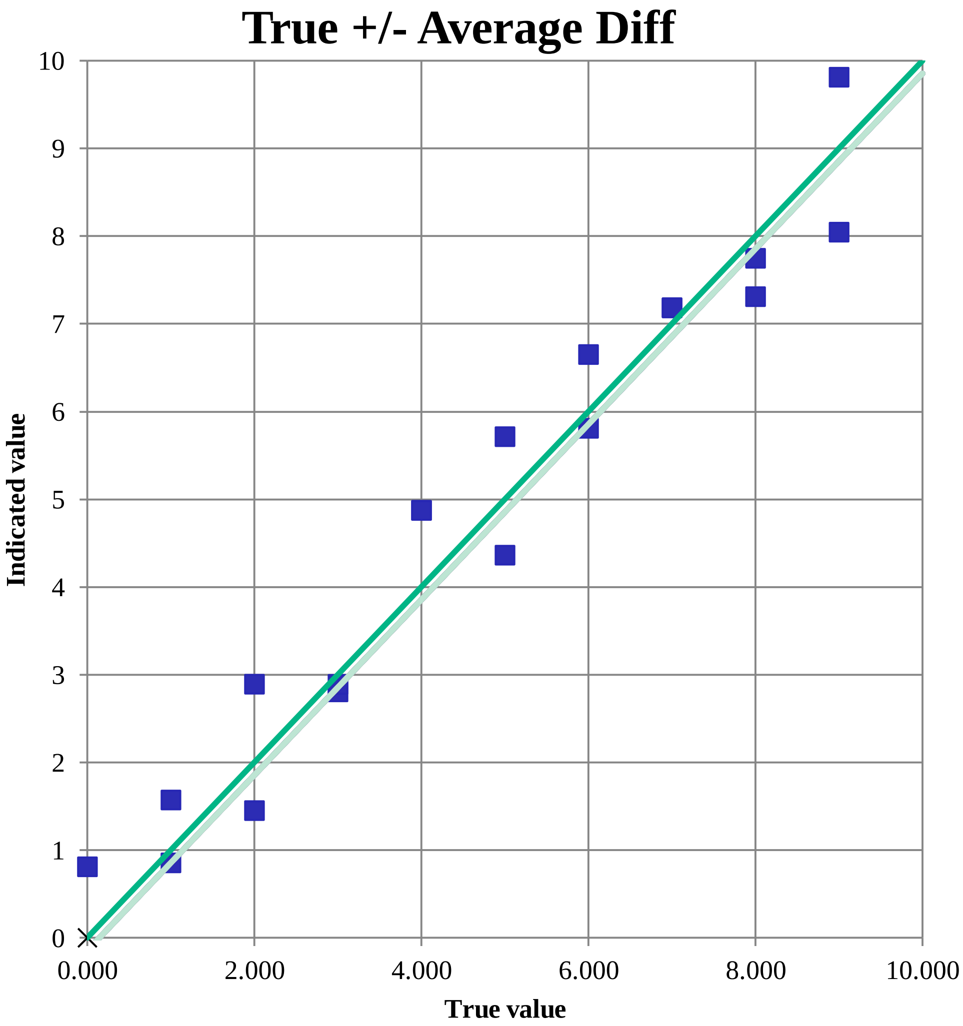


Green line is 45 degree (x=y)
Black line is Linear Least Square fit to data pts

The average difference between True and Indicated = 0.148

A line +/- does not give us any information on data scatter.

Simply summing of +/- values can make the average small, because values cancel.



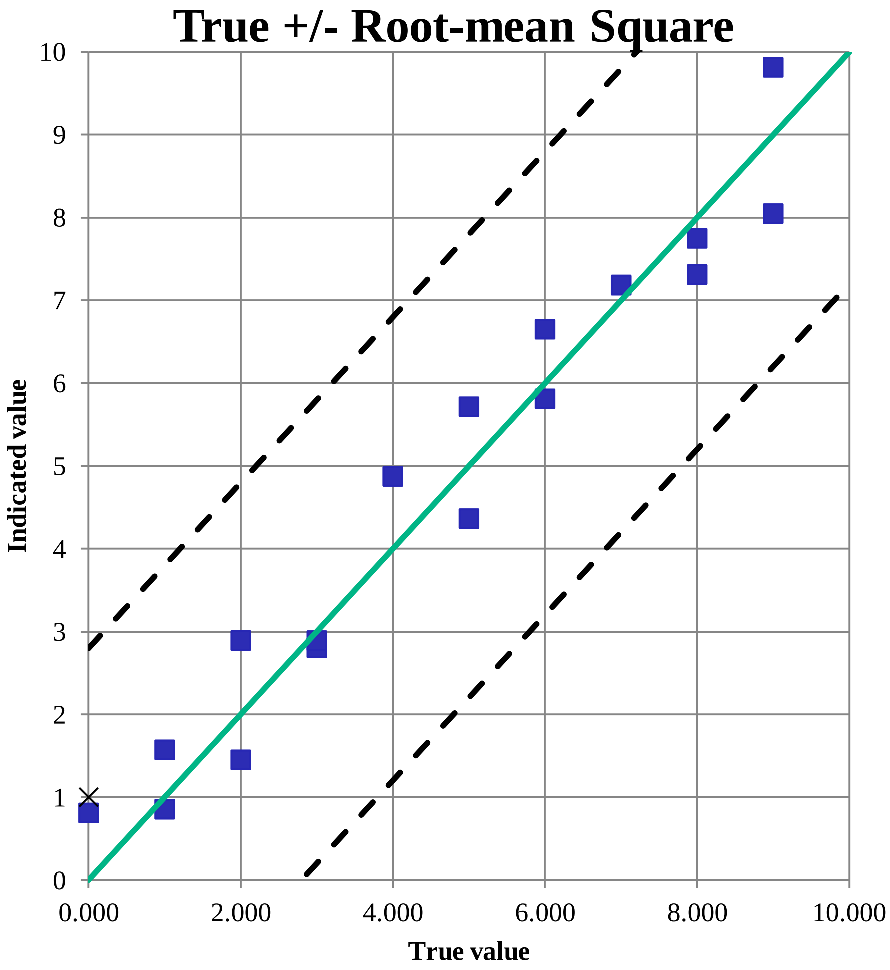
To deal with that, the Root-sum-square estimate can be used:

Measures the deviation of your measurements from theory.

$$RSS = \sqrt{(x_1 - x_a)^2 + (x_2 - x_b)^2 + (x_3 - x_c)^2 \ldots}$$

Where x_i , x_2 ….. x_i = data points
x_a, x_b , ….. = corresponding values in theory

Solves the +/- problem by squaring negative values,
but may also over-estimate the uncertainty.


True +/- Root-mean Square
Indicated value
0
1
2
3
4
5
6
7
8
9
10
0.000
2.000
4.000
6.000
8.000
10.000
True value

Regression (curve fitting):

Computes parameters for your model that minimize the RSS deviation between data and model.

Great if model is simple (straight line). Difficult with complex models.

Note: Excel fits the data with a curve (usually linear) with slope and intercept.

> It also gives you a number labeled R^2
>
> THIS IS NOT THE SLOPE! It is a measure of how good the fit is to the data points, not how good the data itself is.

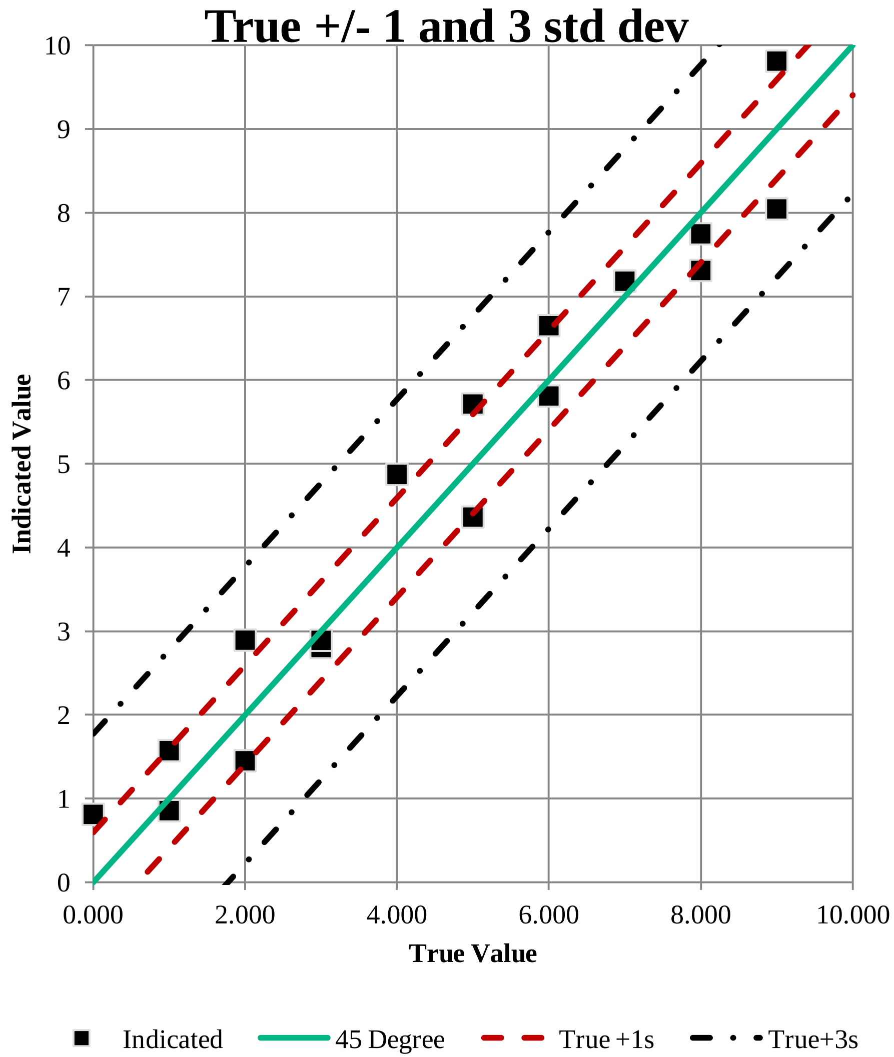
Mean and Standard Deviation of Differences

A classical measure – used and misused:

68% of data is within 1 σ
99% of data is within 3 σ

IF the data scatter is truly random and drawn from a normal (bell shaped) distribution. Few are.

Use, but use with caution.


True +/- 1 and 3 std dev
Indicated Value
10
9
8
7
6
5
4
3
2
1
0
0.000
2.000
4.000
6.000
8.000
10.000
True Value
Indicated
45 Degree
True +1s
True+3s

Complex Models:

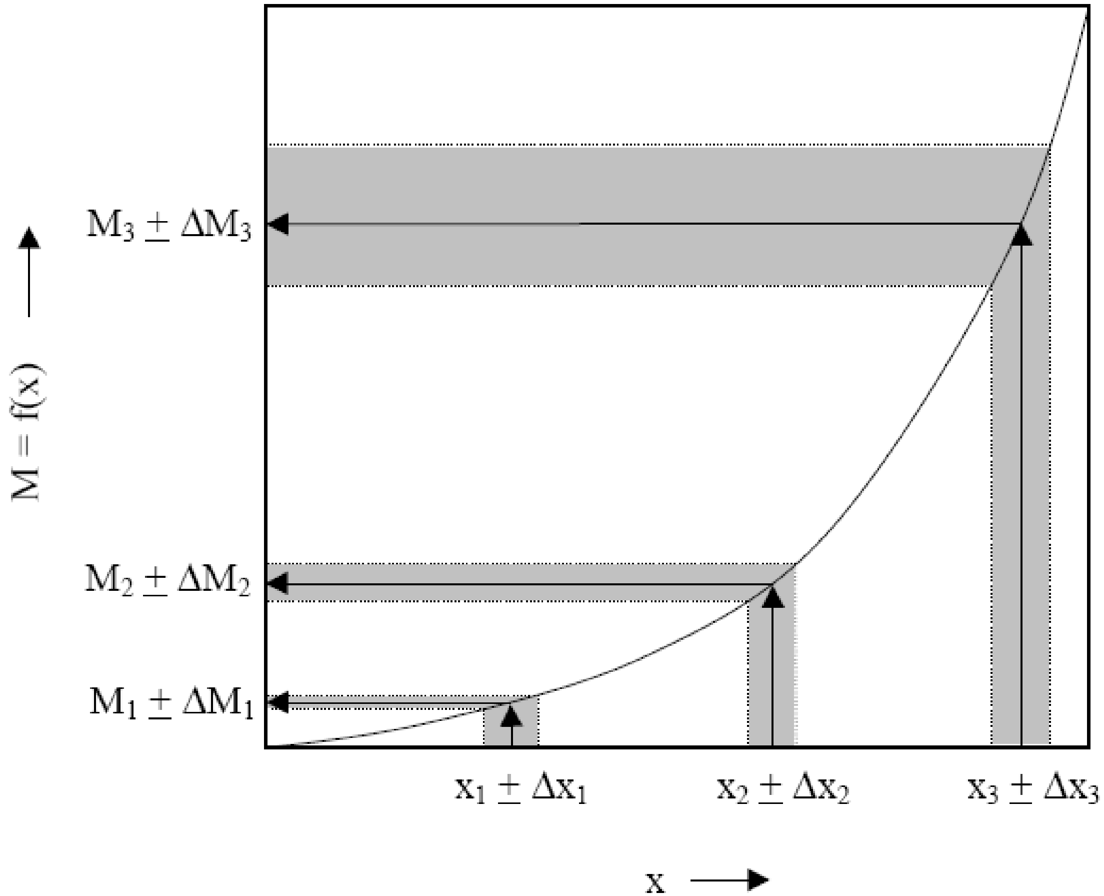
(1) Non-linear models - uncertainty in value changes with conditions.

(2) Quantity being measured is itself a function of other measured parameters. Which ones have the greatest effect?

X1 & X2 were linear systems. X3 has some non-linearity: Force $\approx V^2$

(*A short document posted in Canvas, written by Dr. Prasad several years ago, summarizes how to deal with these. The following is taken from it).*

Non-linear model with one variable:



For given x_i the resulting uncertainly bands ΔM_i are different.

The estimated uncertainty at any point x could be (Taylor series expansion)

$$\Delta M = \frac{\partial f}{\partial x}\Delta x$$

One could write the total uncertainty as an average of:

$$\Delta M = \Delta M_1 + \Delta M_2 + \Delta M_3 \ldots.$$

This would tend to over estimate ΔM and is also not statisically correct.
A better way is the Root-Sum-Square expression,

$$\Delta M = \sqrt{(\Delta M_x)^2 + (\Delta M_y)^2 + (\Delta M_z)^2 + \cdots}.$$

Where the ΔM are from the equation above.

This can be extended to M = f(x, y, z ….) where x,y,z... are parameters, for example pressure, temperature, velocity- each measured with a different instrument. Expansion with Taylor Series, and using the RSS leads to:

$$\Delta M = \sqrt{\left(\frac{\partial f}{\partial x}\Delta x\right)^2 + \left(\frac{\partial f}{\partial y}\Delta y\right)^2 + \left(\frac{\partial f}{\partial z}\Delta z\right)^2 + \cdots}$$

CONCLUSION: What is the purpose of calculating uncertainty?

The customer wants an answer:

X-1 Building residents want water fixtures to work.
No exceptions.

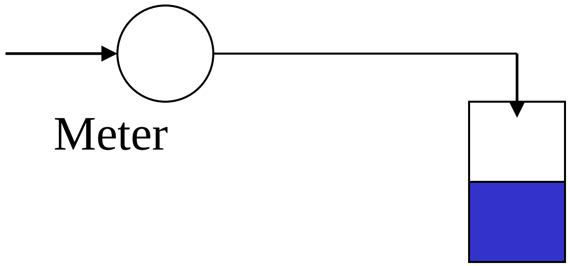
X-2 Process cooling water flow critical, or BOOM!
(+/- 20% uncertainty not good enough)

CONCLUSION: RELATE ALL THIS TO LAB REPORT

• Calculate Uncertainty per Instructions, or Pick one if no instruction

• In Discussion & Conclusions talk about level of uncertainty and possible reasons for it.

A simpler example to illustrate the method: calibrating a flow meter



Bucket with Volume V.
Measure time t to fill.

$Q = V/t \quad m^3 / sec$

$$\begin{aligned}\Delta Q &= \sqrt{\left(\frac{\partial Q}{\partial V}\Delta V\right)^2 + \left(\frac{\partial Q}{\partial t}\Delta t\right)^2} \\ &= \sqrt{\left(\frac{1}{t}\Delta V\right)^2 + \left(\frac{V}{t^2}\Delta t\right)^2}\end{aligned}$$

Or, for expressing uncertainty as a fraction (percent):

$$\frac{\Delta Q}{Q} = \sqrt{\left(\frac{\Delta V}{V}\right)^2 + \left(\frac{\Delta t}{t}\right)^2}$$

Picking some numbers for illustration:

$V = 200$ ml
$\Delta V = +/- 1$ml
$t = 10$ sec
$\Delta t = 0.1$ sec

$$\begin{aligned}\frac{\Delta Q}{Q} &= \sqrt{\left(\frac{1}{200}\right)^2 + \left(\frac{0.1}{10}\right)^2} \\ \frac{\Delta Q}{Q} &= \frac{1}{100}\sqrt{0.25+1} \\ &= 1.1\%\end{aligned}$$

This calculation also tells you which parameter has the most influence.

How do you know what the Δ's are?

That is up to you:

- Sometimes the instrument manufacturer specifies (e.g., % of scale)
- Sometimes observations during data recording suggest
- They can depend on where you are operating
- Sometimes it is your honest opinion and engineering judgment.

It's statistical after all.

Example for Lab X-3:

F = mV = (ρ A V) V = ρ A V^2

Consider sources of uncertainty

$$\Delta M = \sqrt{\left(\frac{\partial f}{\partial x}\Delta x\right)^2 + \left(\frac{\partial f}{\partial y}\Delta y\right)^2 + \left(\frac{\partial f}{\partial z}\Delta z\right)^2 + \cdots}$$

Propagation of Error expression

Uncertainty in the density, ρ.

It is a value out of a table of properties. What can you say about its uncertainty? The question is, how closely can you read the thermometer? And is the reading representative?

Temperature	Density	Spec Heat….
…..	…..	……
19 C	1000	
20	999	
21	998	

$$\Delta\rho/\Delta T = (998-1000)/(21-19) = 1 \text{ kg/m}^3 \text{ per } {}^{o}\text{C}$$

(Values approximate, as an example only)

Area A and velocity V:

Does this jet have a vena contracta? If so, you measured a coefficient of contraction in Lab X2. Uncertainty in the area, ΔA could be on the order of that contraction.

Velocity, V is calculated from the presumed area, so uncertainty in the area propagates to the velocity. The velocity also depended on measurement of time and volume flow.

So the Velocity uncertainty can be expanded using the relation $V = Q/t$ where Q is the measured quantity collected in graduated cylinder in time t.