Ecole polytechnique fédérale de Lausanne EPFL
School of Architecture, Civil and Environmental Engineering ENAC
ENAC - Environmental Science and Engineering Section ENAC-SSIE

Uncertainty in MFA

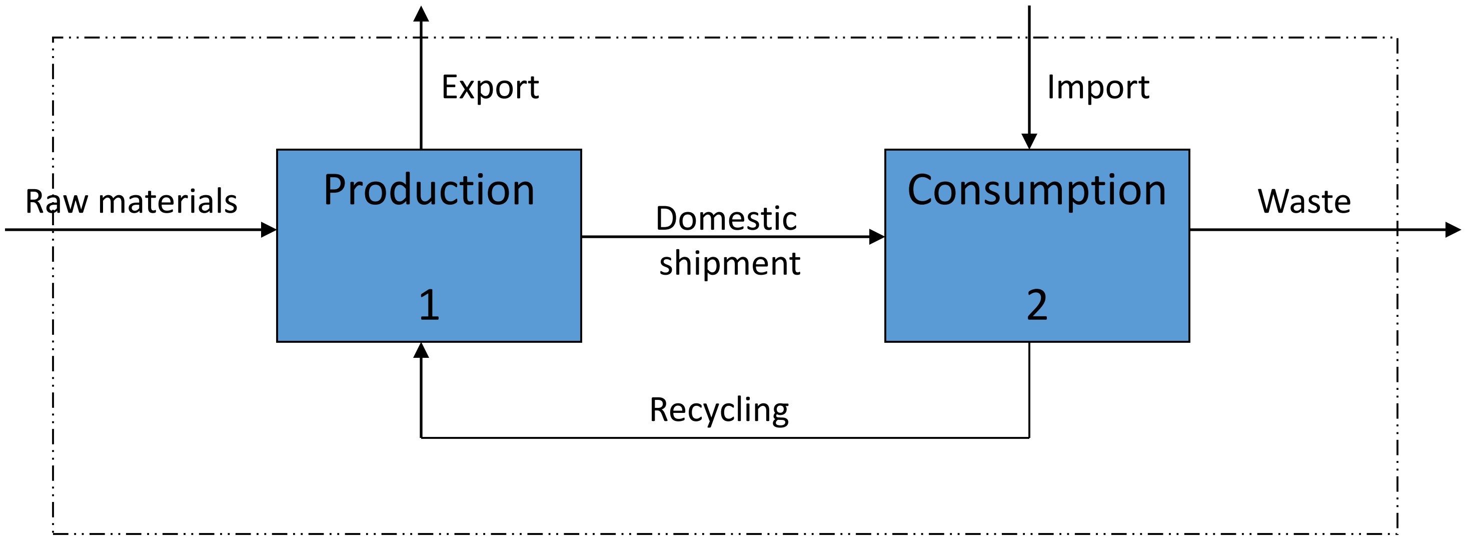
Stefan Pauliuk, Industrial Ecology Group, University of Freiburg, Germany

Background: Solving MFA systems with the analytical system solution

Solving MFA systems – The analytical system solution

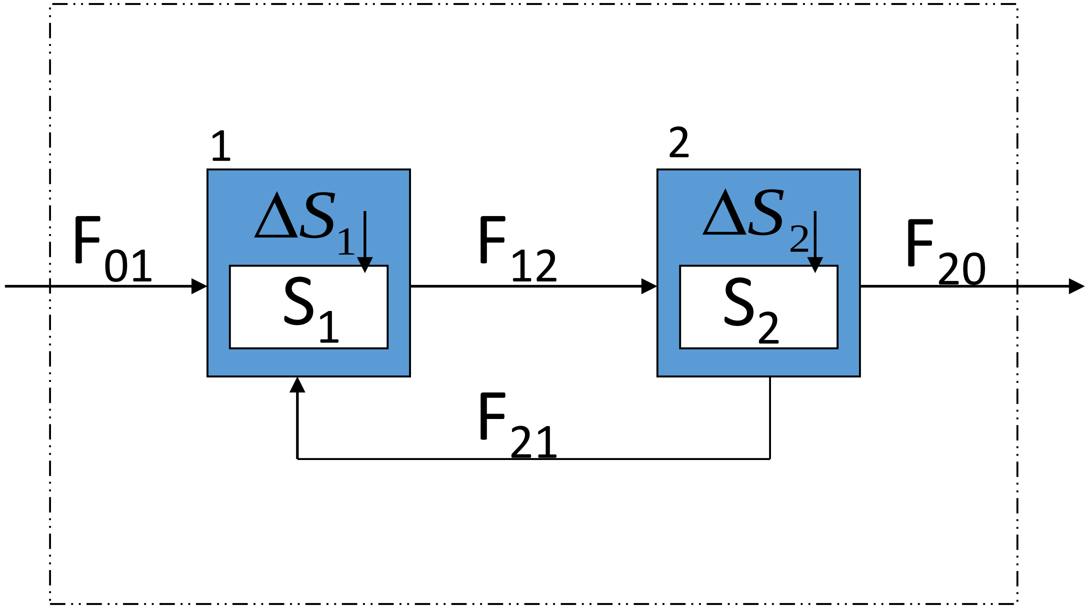
Problem setting: The *recycling rate* in the following system is expected to increase by 10%. Given that domestic shipments are equal, how does the demand for raw materials change?

By less than 10%?
By exactly 10%?
By more than 10%?



Problem designed by Daniel B Müller, NTNU

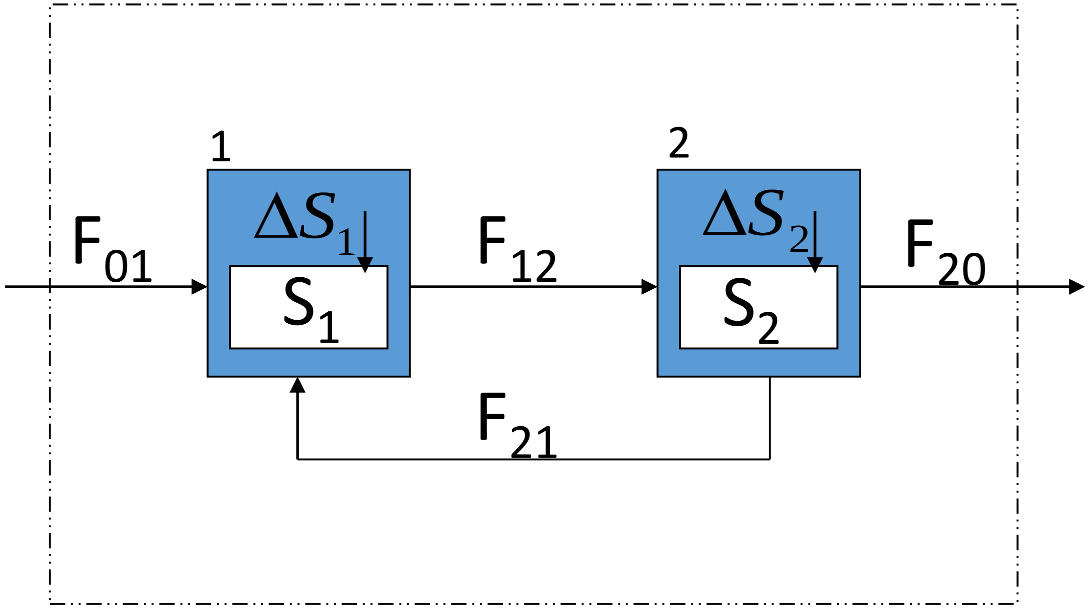
First step: Name all processes and system variables and define the recycling rate



$$\text{Recycling rate } \alpha := \frac{F_{21}}{F_{12}}$$

Graph by Daniel B Müller, NTNU

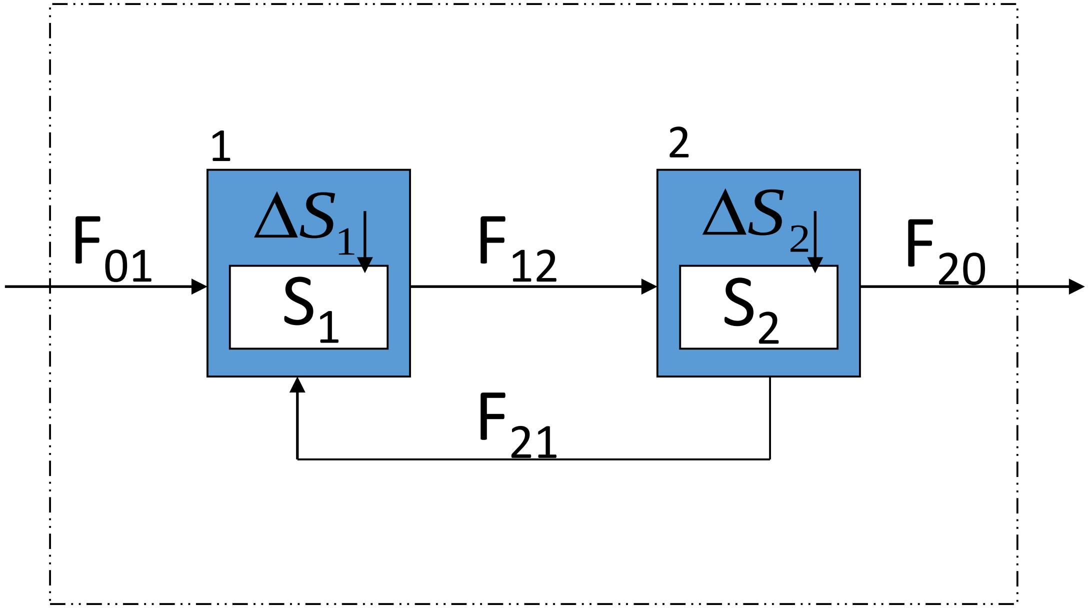
Second step: Determine how much information is needed



Each (linearly independent) equation involving the system variables represents one unit of information.

- We have **eight** system variables (four flows, two stock changes, two stocks)
- The mass balance equations give us **two constraints**
- Hence, we need **six** independent pieces of information (equations) to quantify the system!

Third step: Gather information needed, make simplifications



- Static model:
  We are only interested in the system's throughput performance, and therefore drop the stocks from the list of system variables. (two pieces of information less) (Yes, dropping stocks is OK!)

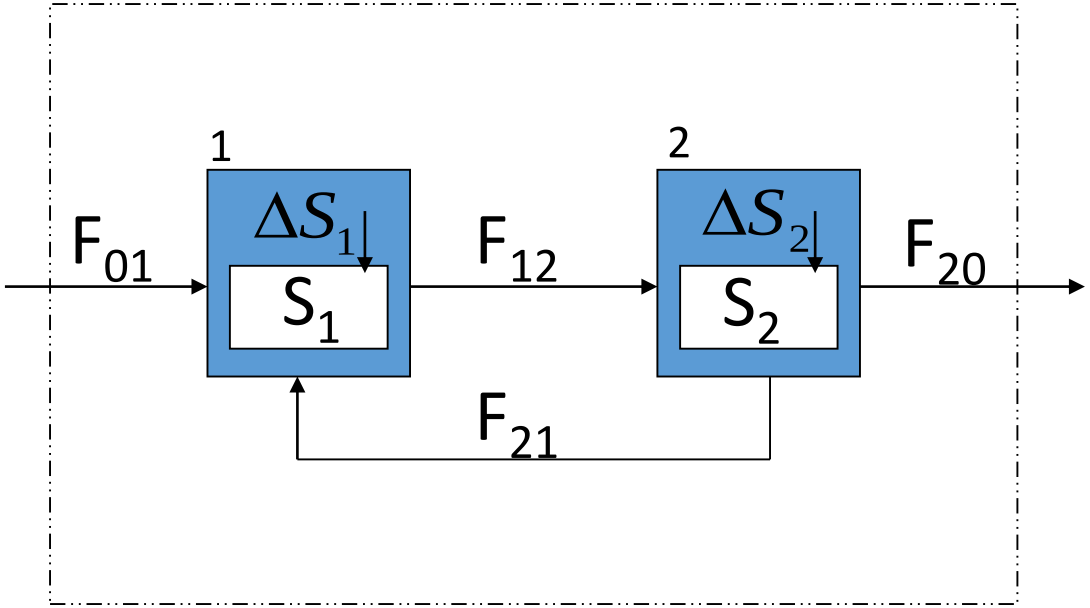
- We assume that throughout the year, there is no net accumulation of material in the processes:
  $\Delta S_1 = 0$, $\Delta S_2 = 0$ (0 is a parameter, two pieces of information)

- We consider the inflow to the system, F_{01}, as given: $F_{01} = D$ (D is a parameter, one piece of information)

- We introduce the recycling rate α as follows: $F_{21} = \alpha \cdot F_{12}$ (α is a parameter, one piece of information)

→ That makes six pieces of information! We now have all we need to solve the system.

Fourth step: Formulate the systems equations



After having dropped the stocks, we need six systems equations.
The systems equations comprise

1) equations that link the model parameters to the system variables

$F_{01} = D$

$F_{21} = \alpha \cdot F_{12}$

$\Delta S_1 = 0$

$\Delta S_2 = 0$

2) balancing equations (sum in = sum out + net stock addition)

$F_{01} + F_{21} = F_{12} + \Delta S_1$

$F_{12} = F_{21} + F_{20} + \Delta S_2$

Fifth step: Solve the ‘equation system of systems equations’

Method a) Algebraic solution, substitution ‘by hand’

i) Simplify mass balancing equations (stock change is 0)

$$F_{01} + F_{21} = F_{12}$$
$$F_{12} = F_{21} + F_{20}$$

ii) Substitute remaining parameters

$$F_{01} + \alpha \cdot F_{12} = F_{12}$$
$$F_{12} = \alpha \cdot F_{12} + F_{20}$$

iii) Solve above equations for F_{12} and F_{20}

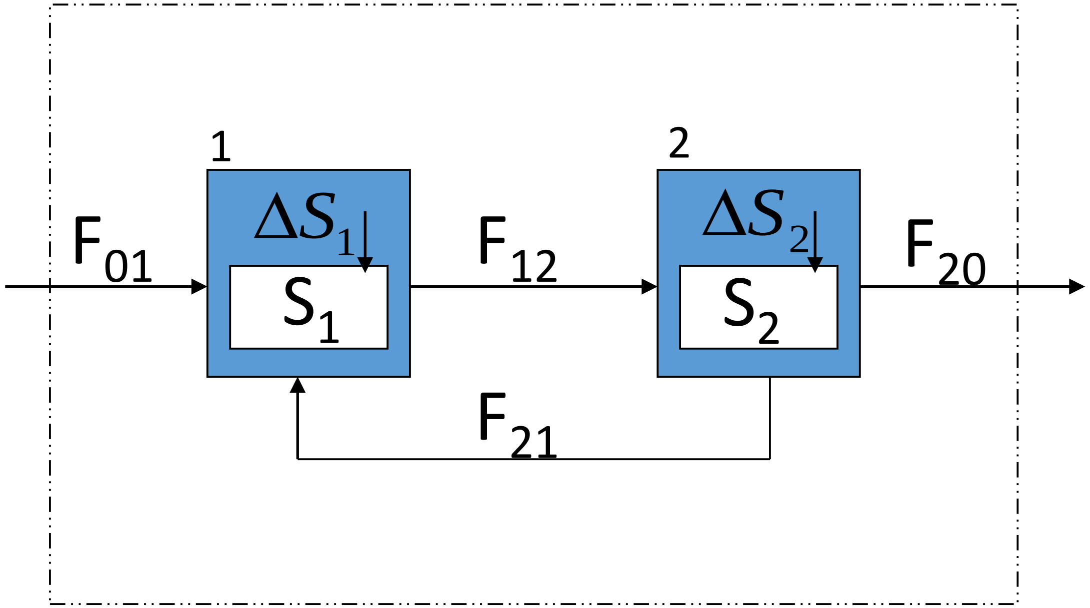
$$F_{12} = D/\,(1 - \alpha)$$
$$F_{20} = D$$

iv) Determine remaining variables

$$F_{21} = \alpha \cdot D/\,(1 - \alpha)$$

v) Perform reality check

There is no outflow other than F_{20}, therefore, F_{20} must equal to $F_{01} = D$.



Final model solution:

$$F_{01} = D$$
$$F_{20} = D$$
$$F_{12} = D/\,(1 - \alpha)$$
$$F_{21} = \alpha \cdot D/\,(1 - \alpha)$$
$$\Delta S_1 = 0$$
$$\Delta S_2 = 0$$

Six system variables (left) equal to
Six functions of the model parameters (right)

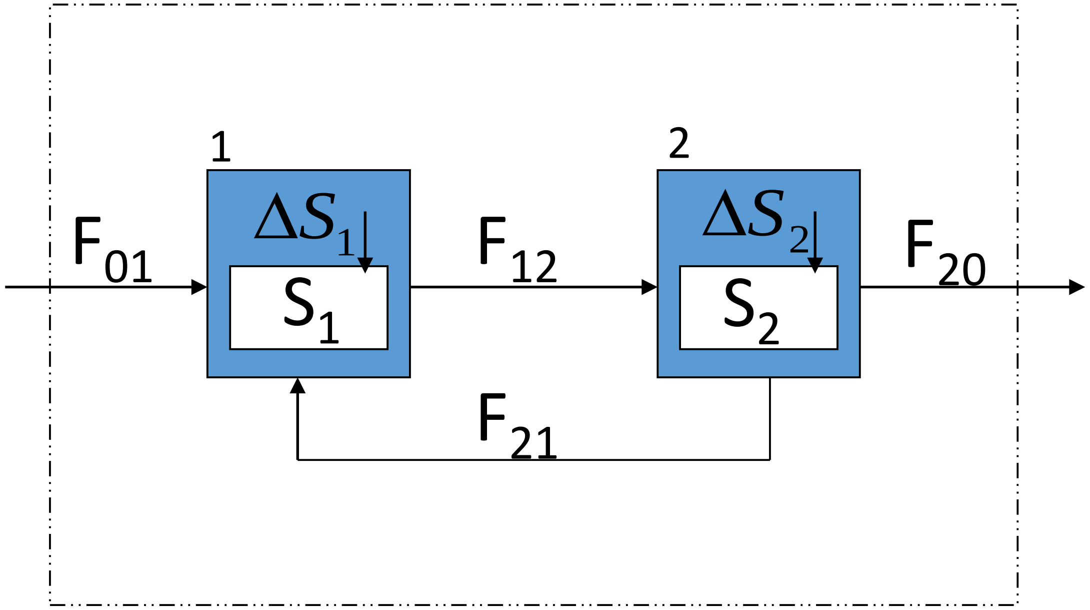
Fifth step professional: Solve the ‘equation system of systems equations’

Method b) Numeric solution, matrix method

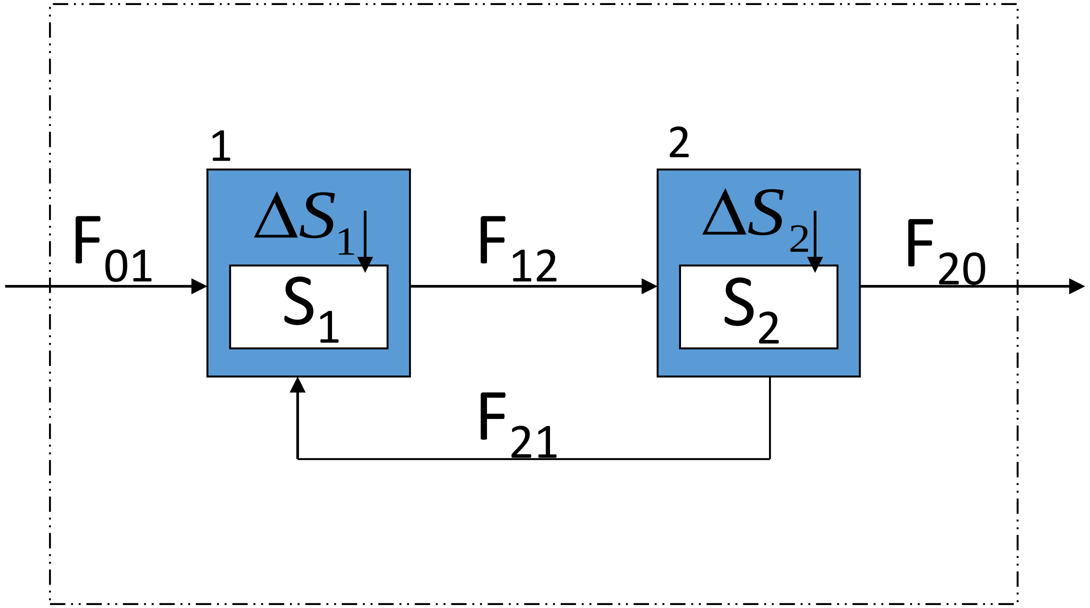
i) Define vector ***x*** of system variables

The order of variables does not matter!

$$x = \begin{pmatrix} \Delta S_1 \\ \Delta S_2 \\ F_{01} \\ F_{12} \\ F_{21} \\ F_{20} \end{pmatrix}$$



Fifth step professional: Solve the 'equation system of systems equations'



Method b) Numeric solution, matrix method

ii) Define linear equation system for ***x***:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{pmatrix} = \begin{pmatrix} A_{11} & \cdots & & & \cdots & A_{16} \\ \cdots & \cdots & & & \cdots & \cdots \\ & & \cdots & \cdots & & \\ & & \cdots & \cdots & & \\ \cdots & \cdots & & & \cdots & \cdots \\ A_{61} & \cdots & & & \cdots & A_{66} \end{pmatrix} \cdot \begin{pmatrix} \Delta S_1 \\ \Delta S_2 \\ F_{01} \\ F_{12} \\ F_{21} \\ F_{20} \end{pmatrix}$$

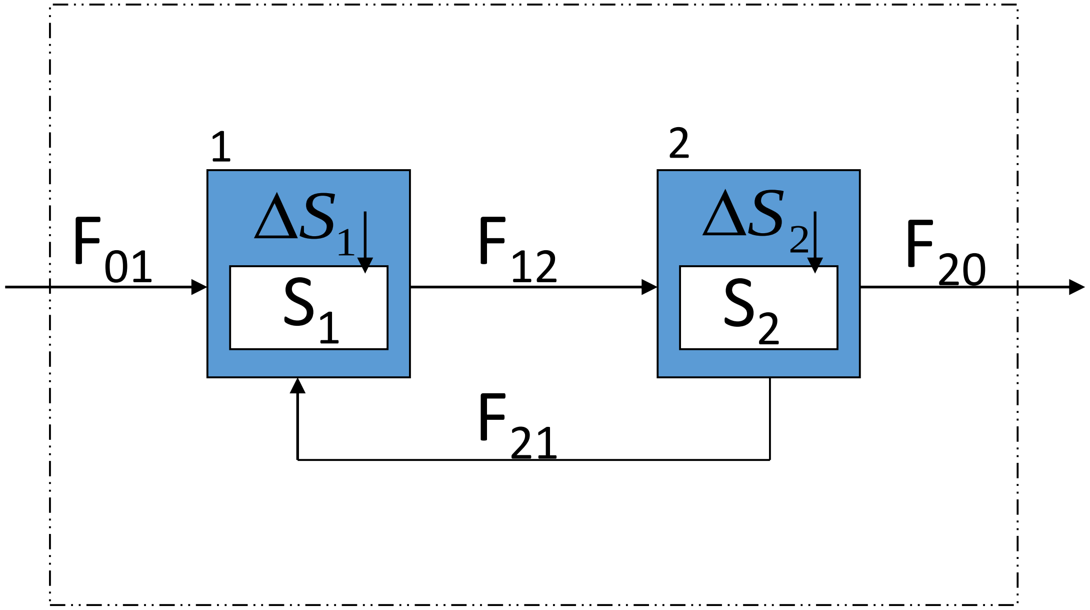
$$y = A \cdot x$$

Help on matrix multiplication: http://www.mathwarehouse.com/algebra/matrix/multiply-matrix.php

Fifth step professional: Solve the ‘equation system of systems equations’

Method b) Numeric solution, matrix method

iii) Translate model equations into equation system



$$\begin{pmatrix} 0 \\ 0 \\ D \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha & -1 & 0 \\ 1 & 0 & -1 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} \Delta S_1 \\ \Delta S_2 \\ F_{01} \\ F_{12} \\ F_{21} \\ F_{20} \end{pmatrix}$$

$$y = A \cdot x$$

The matrix equation is just a different way of writing down an equation system!

The order of equations does not matter!

Model parameters that describe the processes appear in the matrix.

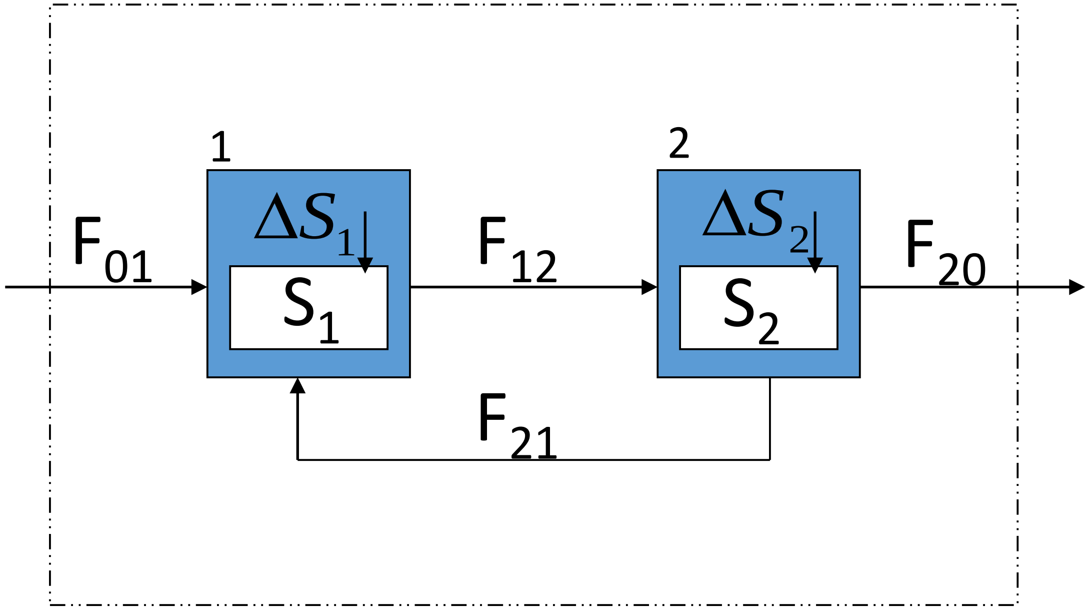
Model parameters that describe exogenous input or output appear in the y-vector.

Help on matrix multiplication: http://www.mathwarehouse.com/algebra/matrix/multiply-matrix.php

Solve the ‘equation system of systems equations’

Method b) Numeric solution, matrix method

iv) Program the matrix into Excel
and solve model equation system for the system variables



$$x = A^{-1} \cdot y$$

Use Excel’s MINV (matrix inversion) and MMULT (matrix multiplication) functions.

For this, you need numeric values for A and α.
Assume that D = 1 and α = 0.25.

(Programs like Matlab or Maple can determine the symbolic matrix inverse.)

Help on matrix multiplication in Excel: https://www.youtube.com/watch?v=G8w-d9U3PJM

Things to remember

1. The system definition dictates which balance equations hold.
   The system definition 'is coded' in the balancing equation (cf. first exercise).

2. Flows always connect to processes or cross the system boundary,
   stocks always are attached to a process.

3. The is no established set of model equations in MFA (unlike LCA and IO),
   and those equations need to fit the problem and data at hand.

4. The matrix method helps to solve large systems with many system variables.

MEFA method

Sensitivity Analysis

Error propagation

Presentation of results

Motivation

The quantitative systems approach has two main advantages compared to qualitative systems understanding:

- **Get the scale right:** (zeroth order effects)
  It tells us what is big and what is small and thus helps to identify ‘real’ problems and solutions and to focus our improvement efforts.
  → You can do this already!

- **Get the direction of change right and acknowledge our ignorance:** (first order effects)
  It tells us which changes are necessary to achieve desired outcomes
  It allows us to study how our ignorance of system structure and data impacts model outcomes and thus our decisions.
  → To learn about how to assess first order effects you need to learn how to deal with model and data uncertainties!

Content:

1) Sensitivity analysis

2) Error propagation:

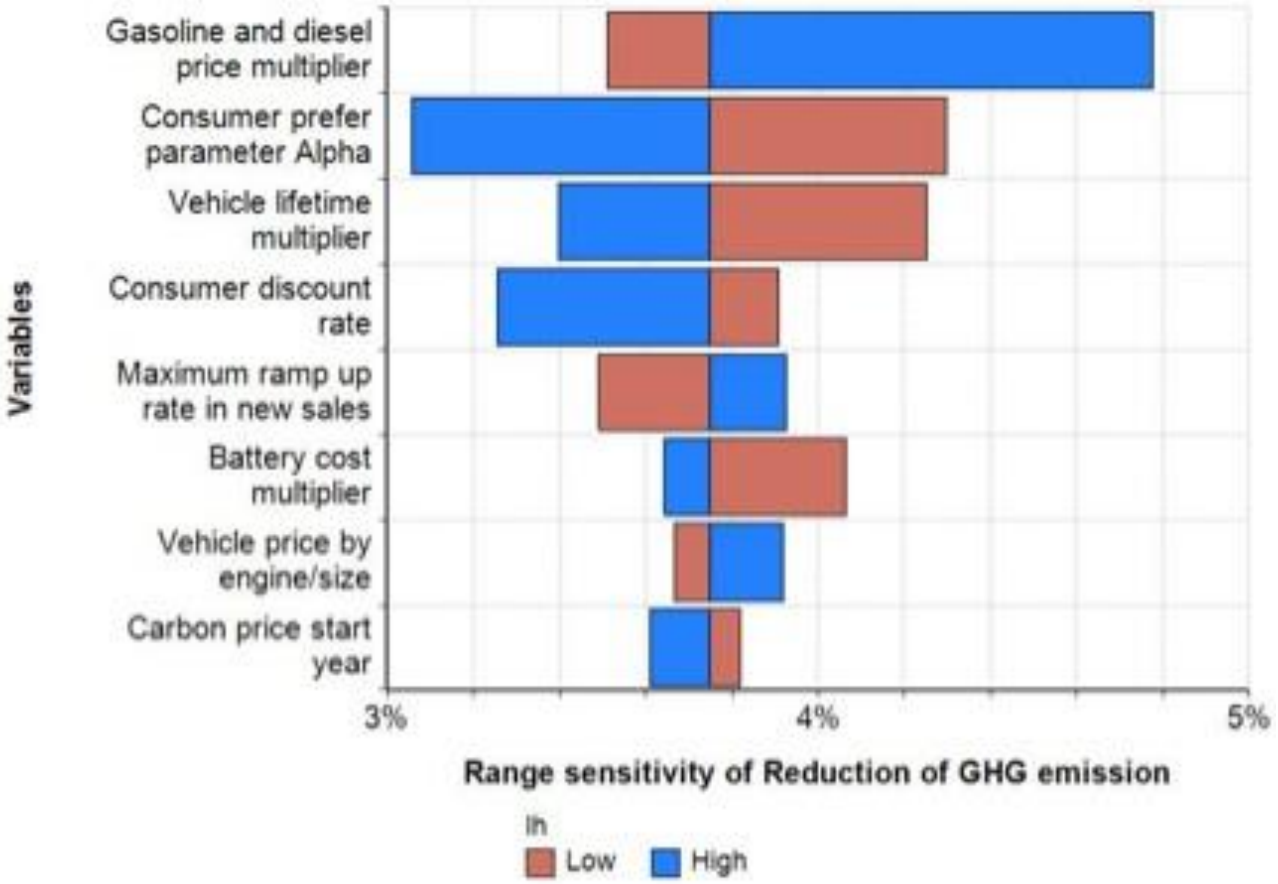
a) Random variables
b) Types of measurement errors
c) How to estimate errors for socioeconomic data?

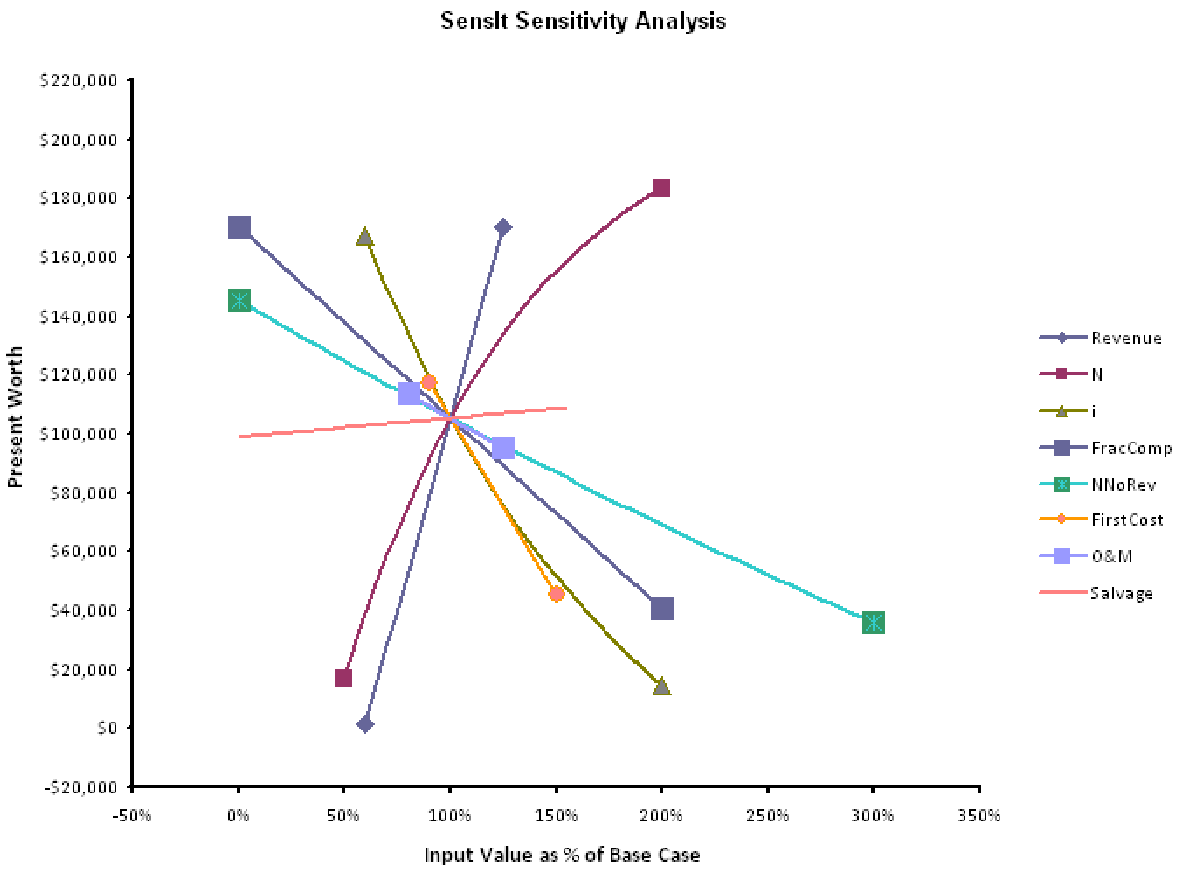
d) How do errors propagate from parameters to system variables?
   i) Analytical approach
   ii) Numerical approach:
      The Monte Carlo Method

3) How to present numerical results and uncertainty ranges

1) Sensitivity analysis …

… is the study of how the output of a mathematical model reacts to changes in input parameters.





Why performing a sensitivity analysis?

- **Better understand your model: study how changes in parameters affect results.**
- **Identify critical parameters and most effective ways to change the system.**
- **Identify critical data points whose uncertainty needs to be reduced.**
- **Compare the response of your model with the expected behavior.**
- **Test for extreme cases**

➔ Sensitivity analysis helps to build **confidence** in your model and the results it produces.

The simplest way of performing a sensitivity analysis: change one parameter at a time

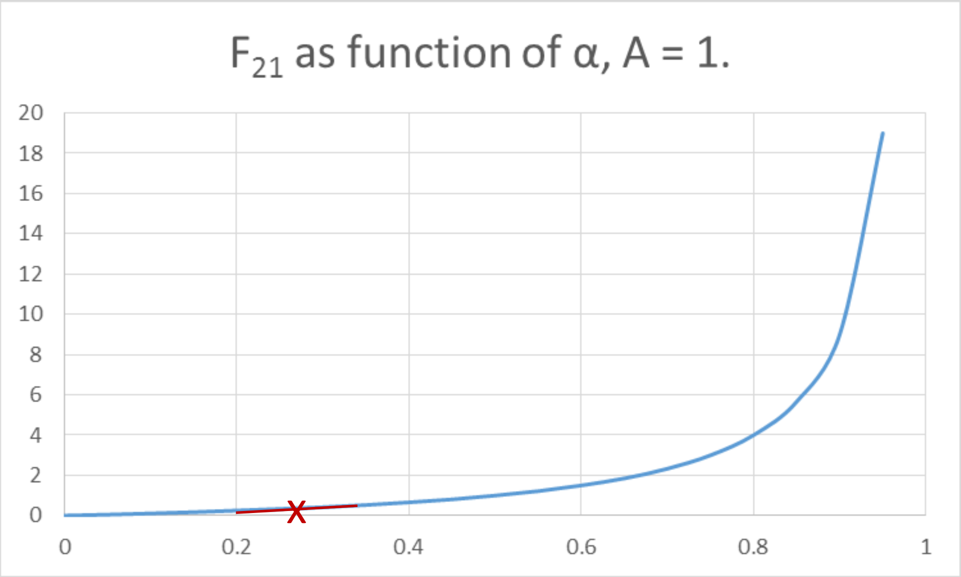
Final model solution (from example in last method lecture):

$F_{01} = A$

$\mathbf{F_{21} = \alpha \cdot A / (1 - \alpha)}$

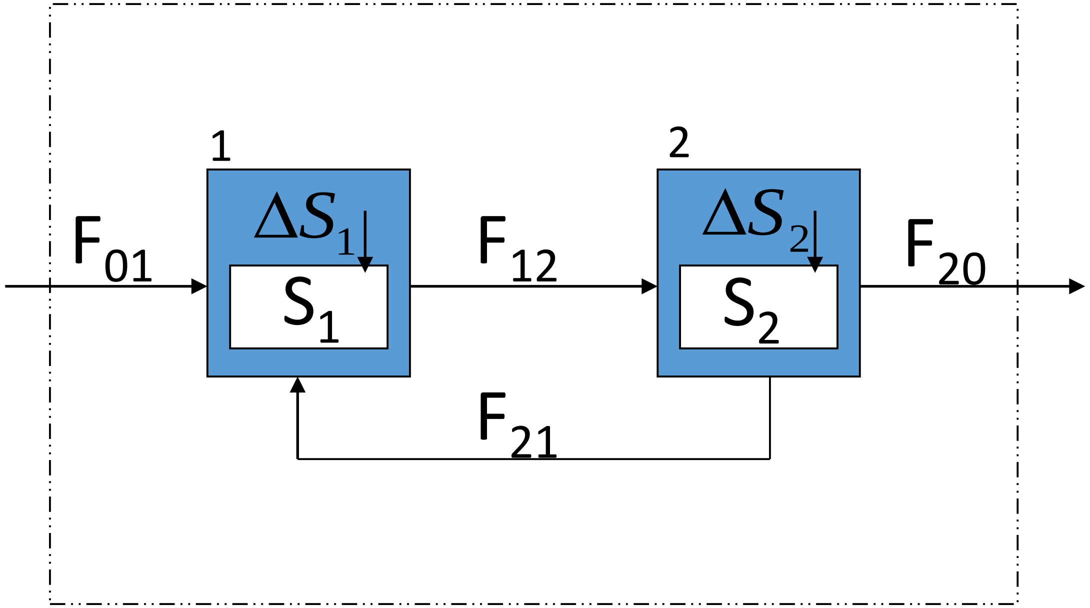
…

Study F_{21} for different values of α:



Base value and slope

α	F_{21}
0	0
0.05	0.05
0.1	0.11
0.15	0.17
0.2	0.25
0.25	**0.33**
0.3	0.42
0.35	0.53
0.4	0.66
0.45	0.81
0.5	1
0.55	1.22
0.6	1.5
0.65	1.85
0.7	2.33
0.75	3
0.8	4
0.85	5.66
0.9	9
0.95	19
1	#DIV/0!



- Most common method to study model behaviour.
- Investigate the effect of small (marginal) changes: Find out how model behaves locally.
- Investigate the effect of alternative parameter choices: scenario analysis (What if?)

Measures of sensitivity

The **absolute sensitivity** of a variable X_i with respect to a parameter p_j, $S(X_i,p_j)$, is the change in X_i per unit of change in p_j. It is the slope of the function $X_i = f(p_j)$. It is calculated as the partial derivative of X_i with respect to p_j.

Absolute sensitivity

$$S(X_i,p_j) = \left.\frac{\partial X_i}{\partial p_j}\right|_{NOP}$$

Use
- calculate change in output due to change in input
- see when a parameter has largest effect on outcome

The **relative sensitivity** of a variable X_i to a parameter p_j, $S(X_i,p_j)$, is the percentage change in X_i per percentage change in p_j.

Relative sensitivity or point elasticity

$$\overline{S}(X_i,p_j) = \left.\frac{\partial X_i}{\partial p_j}\right|_{NOP} * \frac{p_j}{X_i}$$

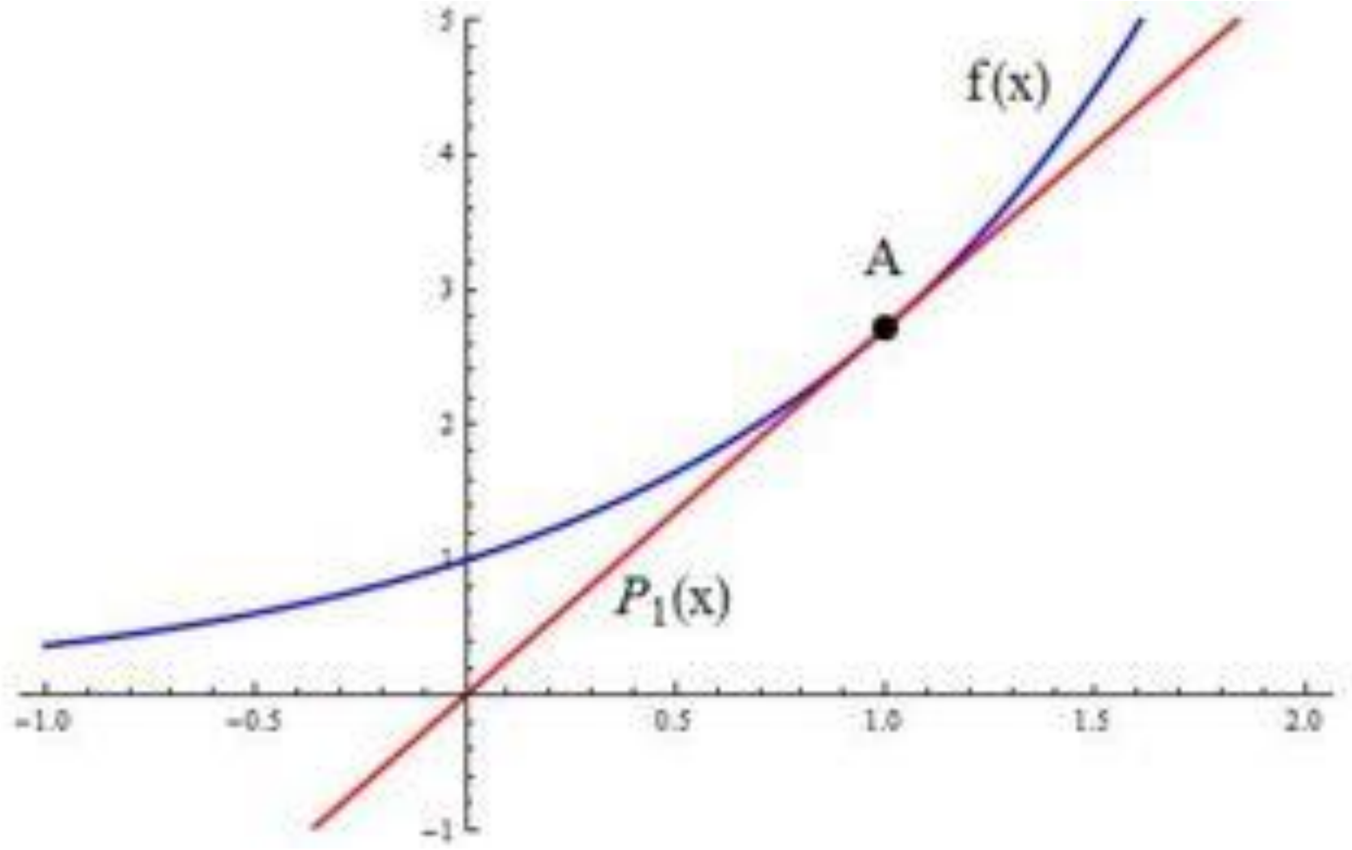
$$\overline{S}(X_i,p_j) = \frac{\%\text{change in } X_i}{\%\text{change in } p_j} = \frac{\Delta X_i / X_i}{\Delta p_j / p_j}$$

Use
- compare impact of parameters on result
- see which parameter has greatest effect

Sensitivities are determined at the normal operating point (NOP), which is the set of initial values for X_i and p_j.

Source: Lecture slide Daniel B Müller, NTNU, modified.

Sensitivity as linear approximation of f(x) by a first-order Taylor series expansion about the point $x_0 = 1$



$$X \approx f(\mu_p) + \left.\frac{\partial f}{\partial p}\right|_{p=\mu_p} * (p - \mu_p)$$

Zeroth and first order of the Taylor series for X

https://de.wikipedia.org/wiki/Taylorreihe

http://mathinsight.org/partial_derivative_examples

Example of sensitivity indicators

From our example above: $F_{12} = A/(1-\alpha)$, $A = 1$, $\alpha = 0.25$. $F_{12}|_{NOP} = 1.33$

Sensitivity of F_{12}	Regarding A	Regarding α
Absolute	$S(F_{12},A)=\frac{\partial F_{12}}{\partial A}=\frac{1}{1-\alpha}\Big\|_{NOP}=1.33$	$S(F_{12},\alpha)=\frac{\partial F_{12}}{\partial \alpha}=\frac{A}{(1-\alpha)^2}\Big\|_{NOP}=1.78$
Relative	$\bar{S}(F_{12},A)=\frac{\partial F_{12}}{\partial A}\cdot\frac{A}{F_{12}}=1$	$\bar{S}(F_{12},\alpha)=\frac{\delta F_{12}}{\delta\alpha}\cdot\frac{\alpha}{F_{12}}=\frac{\alpha}{(1-\alpha)}\Big\|_{NOP}=0.33$

Interpretation of absolute sensitivity:

If p changes by Δp, X changes by S(X,p)·Δp

Interpretation of relative sensitivity:

If p changes by Δ%, X changes by ($\bar{S}$(X,p)·Δ%) %.

Tips for derivative calculation

http://mathinsight.org/partial_derivative_examples

http://library.wolfram.com/webMathematica/Education/WalkD.jsp http://www.numberempire.com/derivatives.php

2) Error propagation

a) Random variables
b) Types of measurement errors
c) How to estimate errors for
socioeconomic data?

d) How do errors propagate from parameters to system variables?
i) Analytical approach
ii) Numerical approach:
The Monte Carlo Method

a) What are random variables?

+ A random variable is a variable whose value varies due to chance.

+ Random variables can take on different values, each with an associated probability.

+ Probability is the measure of the likelihood that an event will occur.

Examples for real-world phenomena commonly modelled as random variables:

+ number shown when throwing dice.

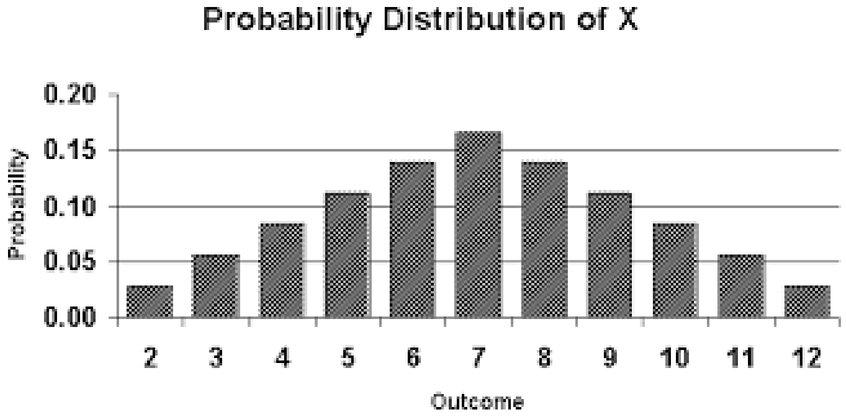
In socio-ecological systems:

+ Content of ingredients or toxic substances in food.
+ Speed and type of next car coming on a road.
+ Income or political opinion of people.

Discrete and continuous random variables

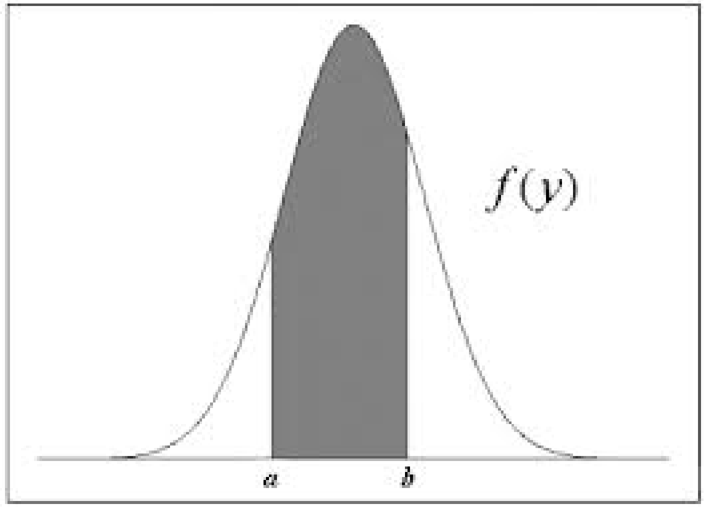
A discrete random variable (*diskrete Zufallsgröße*) assumes only discrete values.

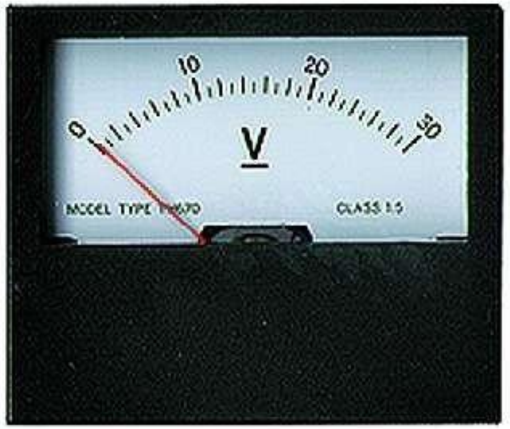
Examples:
+ results of two dice thrown (right)
+ number of people in passenger cars



A continuous random variable (*stetige oder kontinuierliche Zufallsgröße*) assumes values from a continuous spectrum.

Examples:
+ gold content of a sample of mobile phones
+ body height of a sample of people

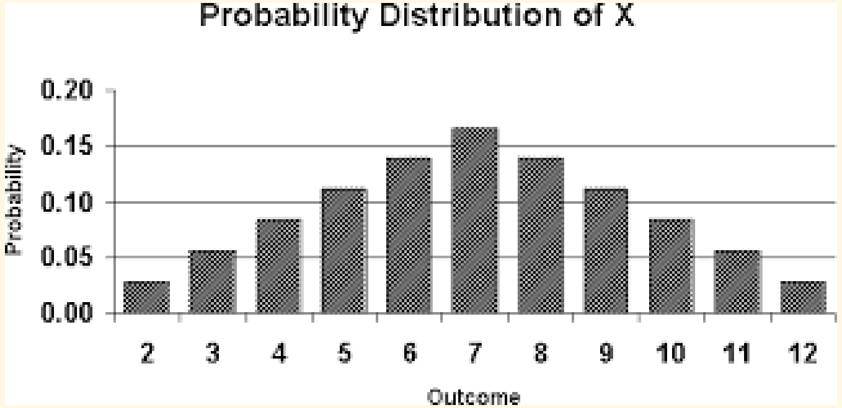
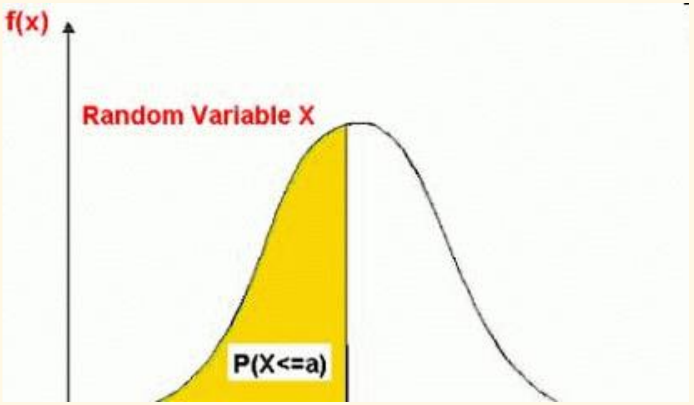
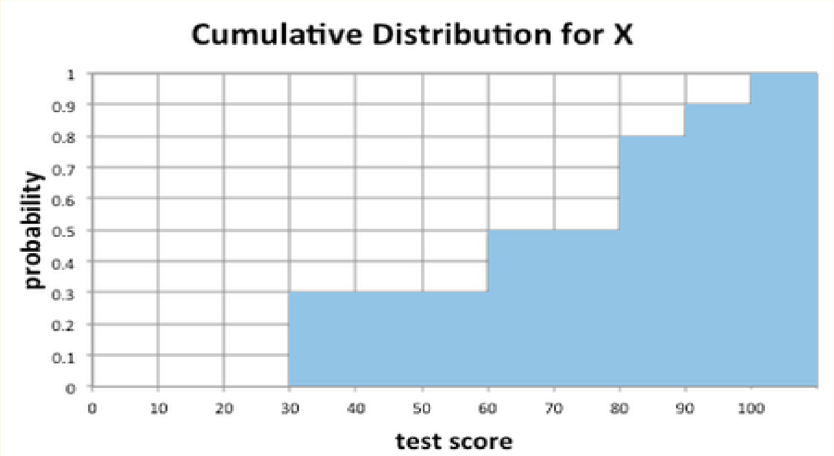
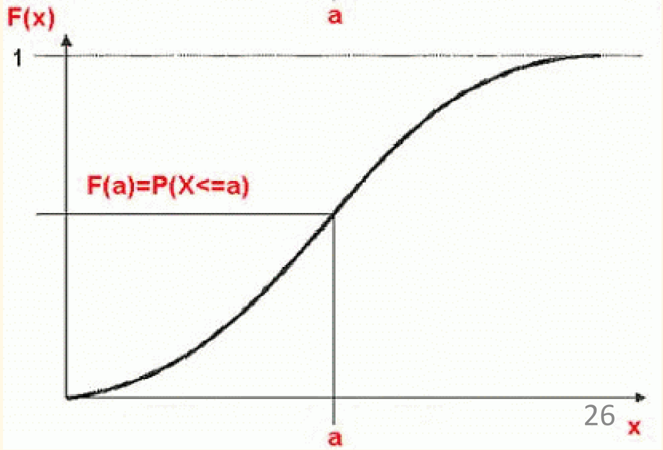




Probability distribution and cumulative probability distribution

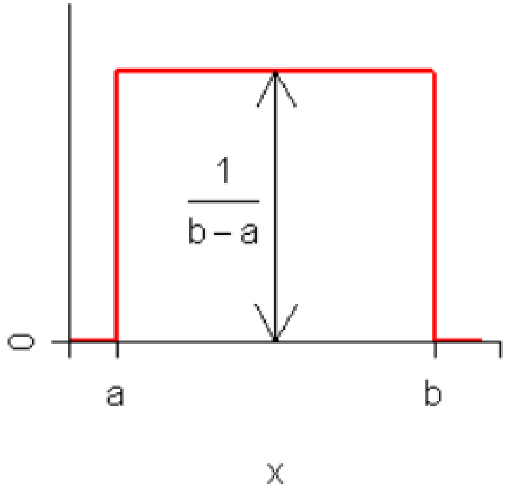
A **probability distribution** is a table of values showing the probabilities of various outcomes of an experiment.

The **cumulative distribution function** (**CDF**) of a random variable *Y*, evaluated at *x*, is the probability that *Y* will take a value less than or equal to *x*.

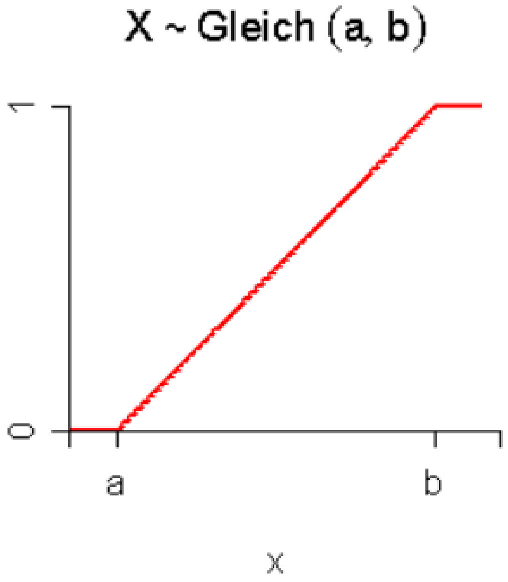
	Discrete	Continuous
Probability distribution (function) **f(x)** **PDF ≠**		
Cumulative probability distribution function **F(x)** **CDF**		

Example for continuous probability distribution: Continuous uniform distribution *(stetige Gleichverteilung)*

$$f(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{sonst} \end{cases}$$



$$F(x) = \begin{cases} 0 & x \le a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \ge b \end{cases}$$



Can be used as conservative estimate when it is clear that a random variable has definite lower and upper boundaries but little is know about its distribution.

Example:

Carbon content of steel:

Cannot be lower than 0

Cannot be larger than 2.1 % (def. of steel)

In Excel:

Function Random/*Zufallszahl* returns a random number with a = 0 and b = 1.

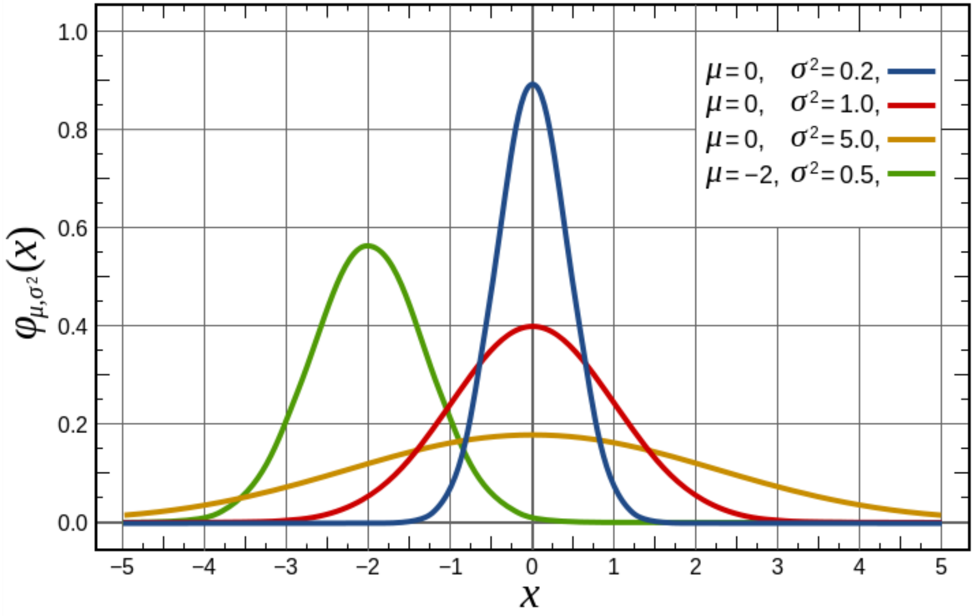
= Random()

= Zufallszahl()

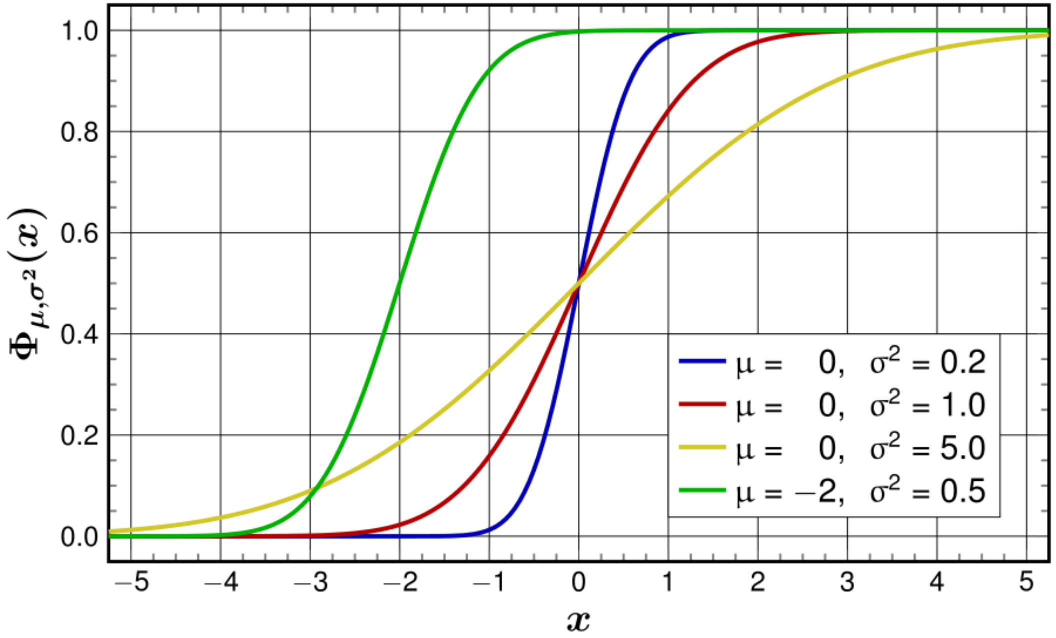
https://de.wikipedia.org/wiki/Stetige_Gleichverteilung

Example for continuous probability distribution: Normal distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} \mathrm{d}t$$



In many natural and technical processes, random errors follow a normal distribution.
Superposition of many independent random variables leads to a normal distribution (central limit theorem)
The normal distribution has two parameters:

μ: indicates the maximum value and also the expected value of a large sample
σ: indicates the spread of the distribution

In Excel: =NORM.INV(RANDOM();μ;σ) generates a normally distributed number with parameters μ and σ.

https://de.wikipedia.org/wiki/Normalverteilung

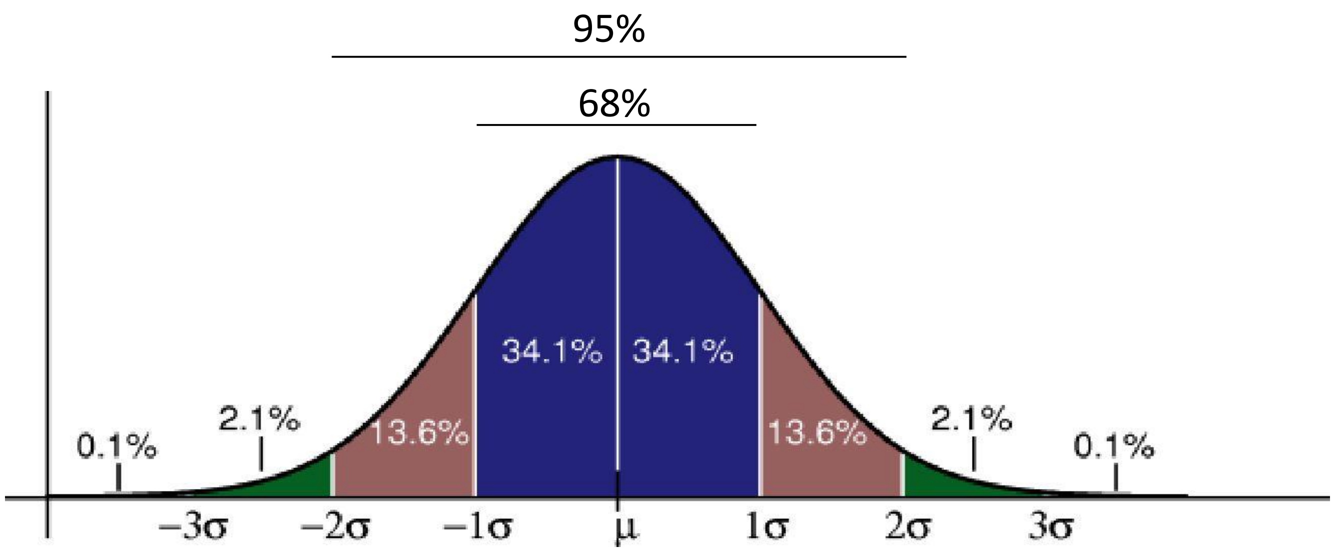
The confidence interval

A confidence interval is an observed range that frequently contains an unobservable parameter
if the experiment is repeated.
The frequency is determined by the confidence level.

We *say* that “We are 99% confident that the true value of the parameter is in our confidence interval.”

By that we *mean* that “99% of the observed confidence intervals, if the experiment is repeated,
will contain the true value”.

Normal (Gaussian) distribution of measurements



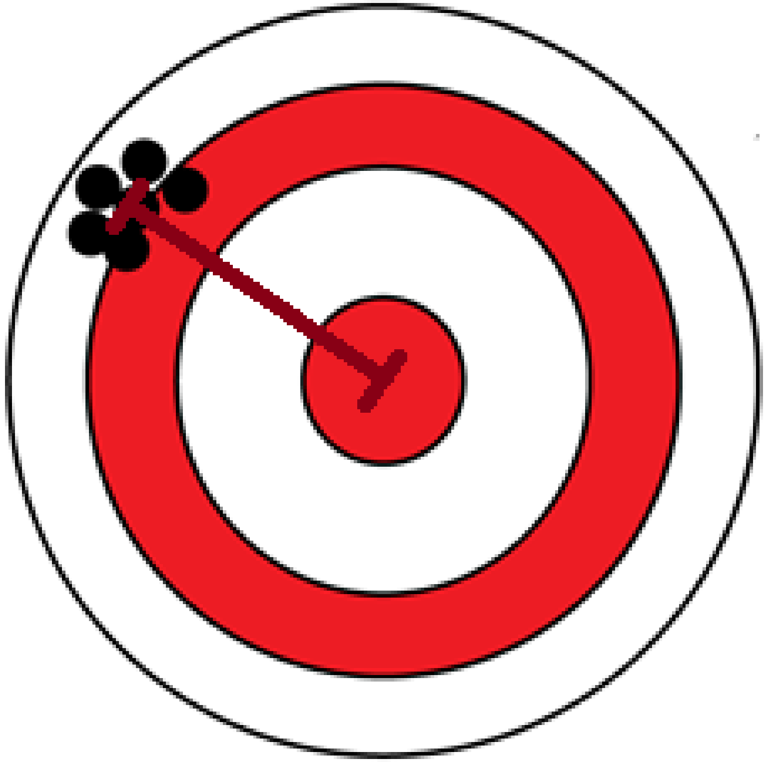
http://introcs.cs.princeton.edu/java/11gaussian/images/stddev.png

b) Systematic and random measurement errors

Definition of measurement error:

Measured value = reference ('true' value) + measurement error

Measurement error = Systematic error + random error



Characteristics of systematic errors *(systematische Fehler)*

→ Consistency: All measurements are systematically too high or too low compared to reference value

→ Cannot be quantified if experiment is repeated under the same conditions (random variable!), alternative measurement methods are needed!

→ The causes of systematic errors can in principle be identified

example: + effect of air on speed of light not considered in distance measurement
+ boiling point variation with altitude not considered in temperature measurement
+ measurement device not correctly calibrated (thermometer, ruler, ampere-meter)

Systematic and random measurement errors

Definition of measurement error:

Measured value = reference ('true' value) + measurement error

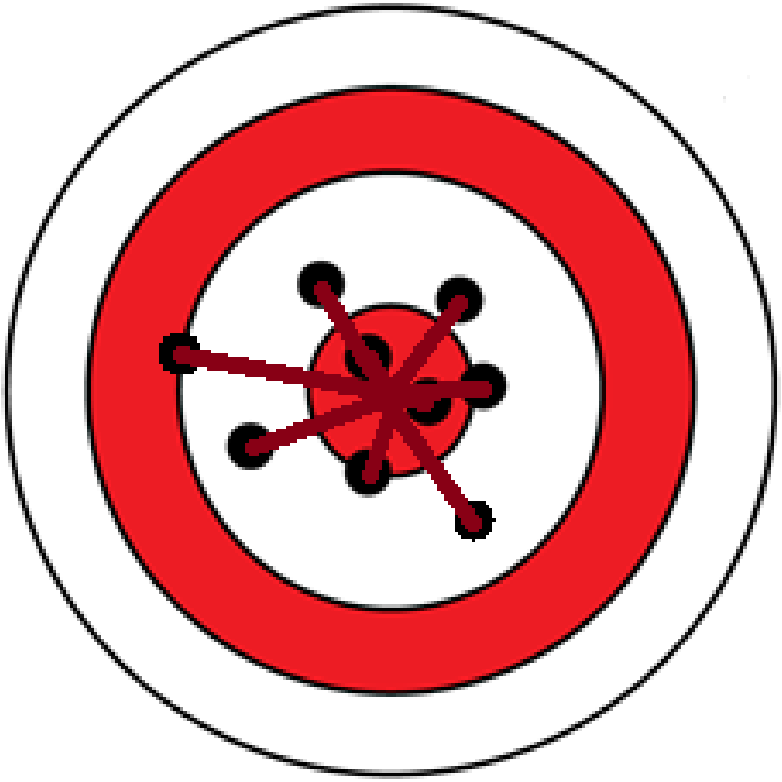
Measurement error = Systematic error + random error

Characteristics of Random errors *(zufällige Fehler)*

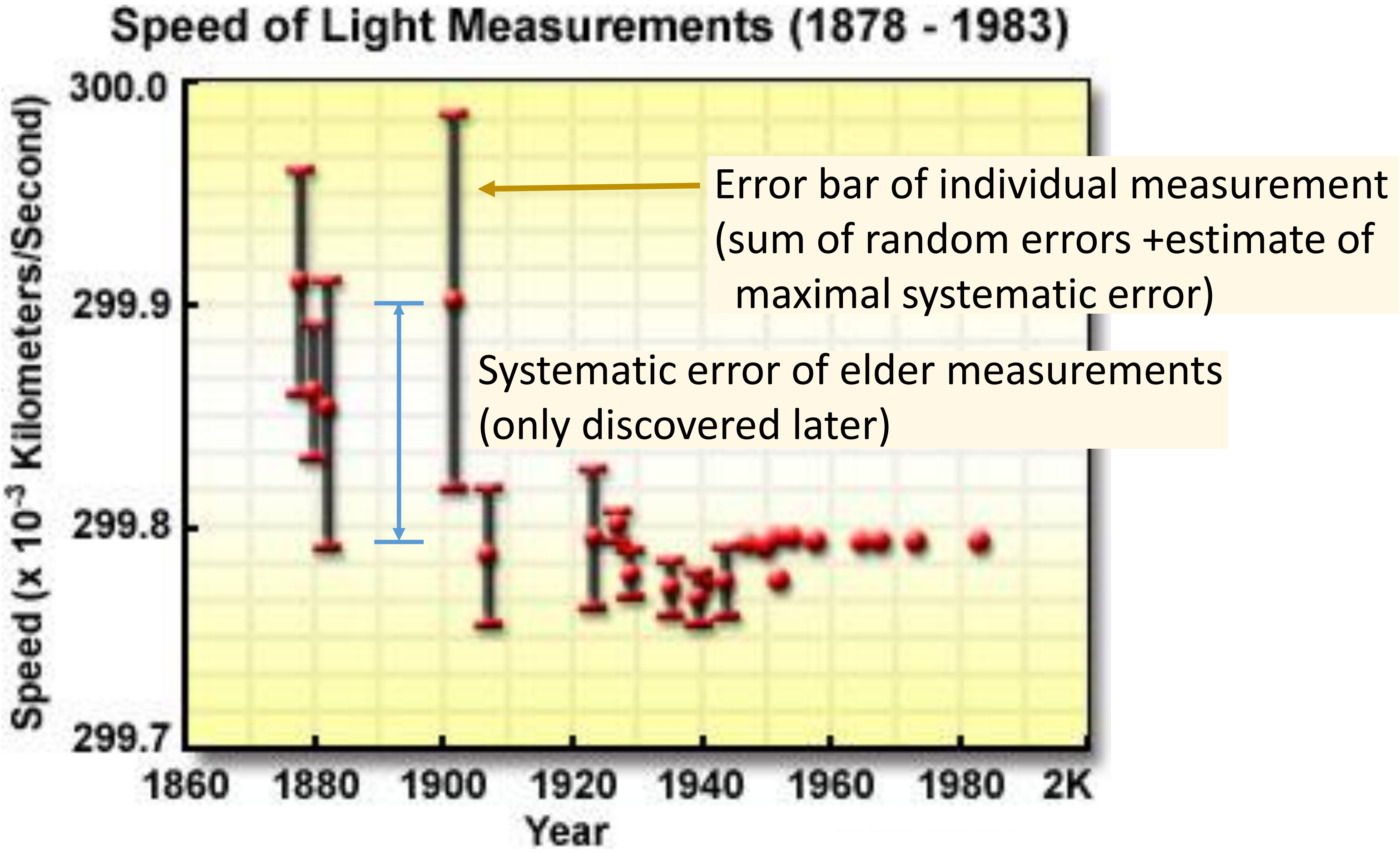
→ Inconsistency: random errors vary in sign and magnitud

→ Random errors are a result of fluctuations in measured object and measurement device

→ With increasing sample size the probability that randor errors cancel out across measurements increases.



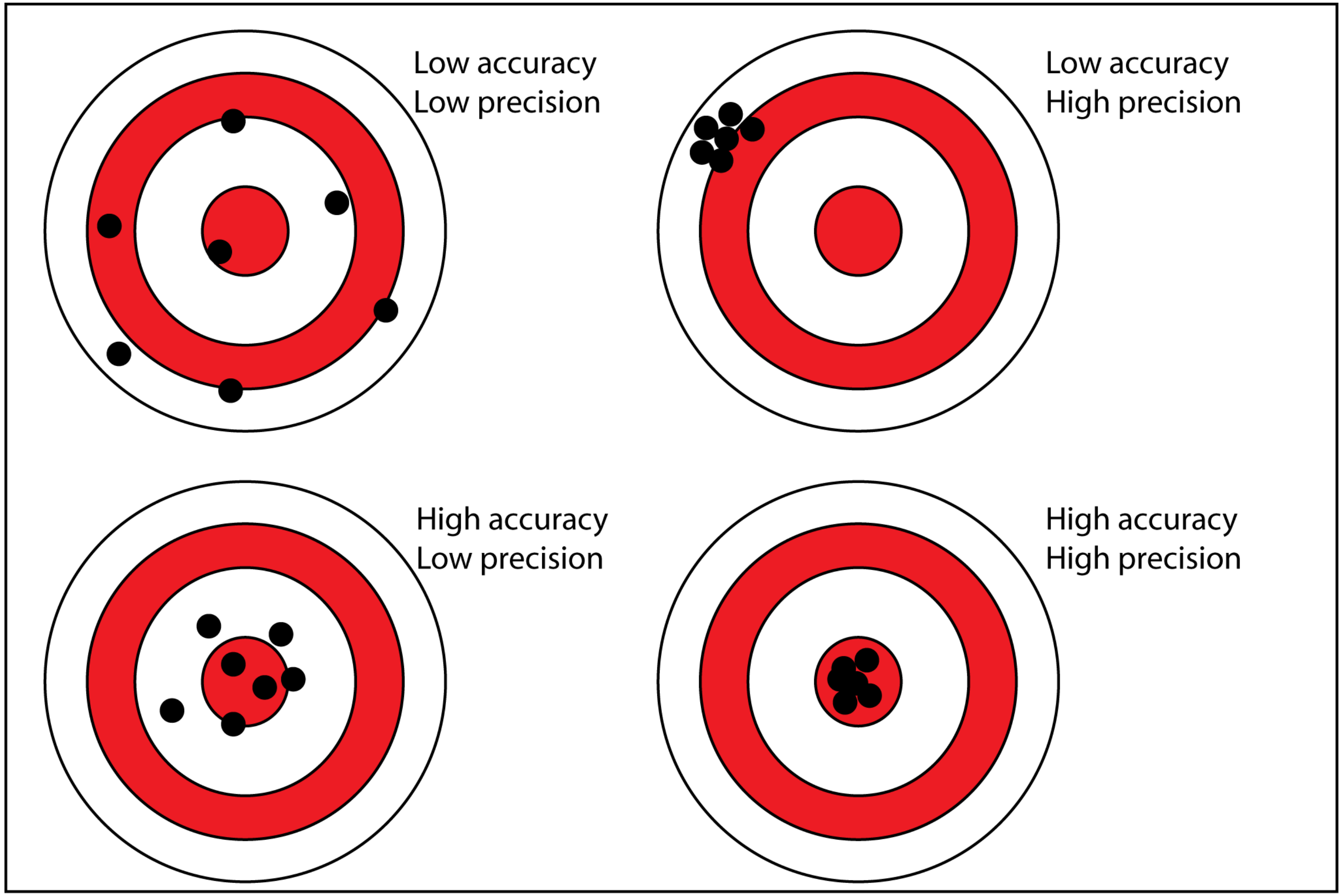
Example: The history of measuring the speed of light



http://olympus.magnet.fsu.edu/primer/lightandcolor/speedoflight.html

Accuracy vs. precision
Richtigkeit und Präzision

Der Alltagsbegriff ‚Genauigkeit' wird in der Meßtechnik in zwei getrennte Konzepte zergliedert:
Richtigkeit (Abweichung des Mittelwertes der Meßreihe vom Referenzwert)
und Präzision (Streubreite der Meßreihe).



Accuracy ←→ Systematic error

Precision ←→ Random error

c) How to deal with measurement errors for socioeconomic data

General problem: Often, no multiple measurements are available.

- Government statistics
- Industry statistics
- Expert estimates
- Secondary sources (newspaper, blogs, …)

The measurement error therefore cannot be properly identified and has to be estimated.

How to estimate systematic errors of socioeconomic data:

- **Need to know the context of the measurement!**
  - Are the data derived from a mandatory scheme? (e.g., car registration, hazardous waste)
    → Indicates low systematic error
  - Are there incentives to over- or underreport (emissions, revenue)
    → Indicates direction of systematic error
  - Is there any balance check that can be done?
  - Do the data represent a single industrial plant or sectoral average?
  - Look for records that document under- or over-reporting

d) How do errors propagate from parameters to system variables?

i) Analytical approach

If the error estimates Δp, Δq, Δr, … of the model parameters p, q, r, … are given, their impact on the result X = f(p,q,r, …) can be calculated as follows:

$$\Delta X = f(p+\Delta p, q+\Delta q, r+\Delta r, \ldots) - f(p,q,r,\ldots)$$

for small variations: dp = Δp, dq = Δq, dr = Δr, we can apply a Taylor series expansion of dX = ΔX and find

$$dX = \left(\frac{\partial f}{\partial p}\right)dp + \left(\frac{\partial f}{\partial q}\right)dq + \left(\frac{\partial f}{\partial r}\right)dr + \ldots$$

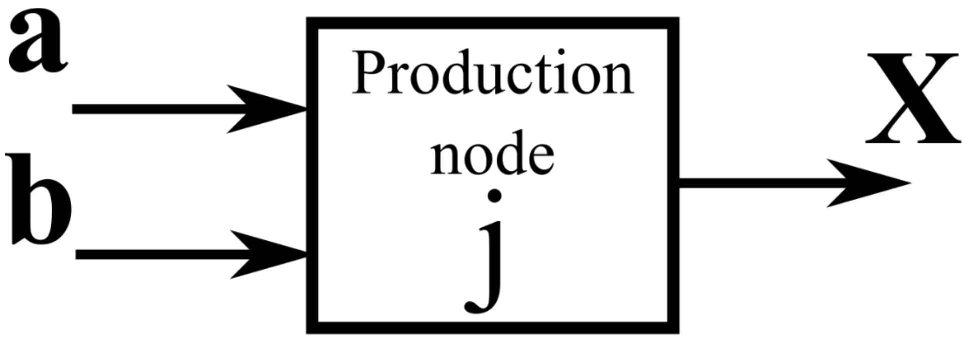
dX is the total differential of X. It gives the change in X as a result of small variations in the parameters dp, dq, dr, … For small deviations we estimate the maximum possible error from a situation in which the errors of the parameters p, q, r, … all point in the same direction:

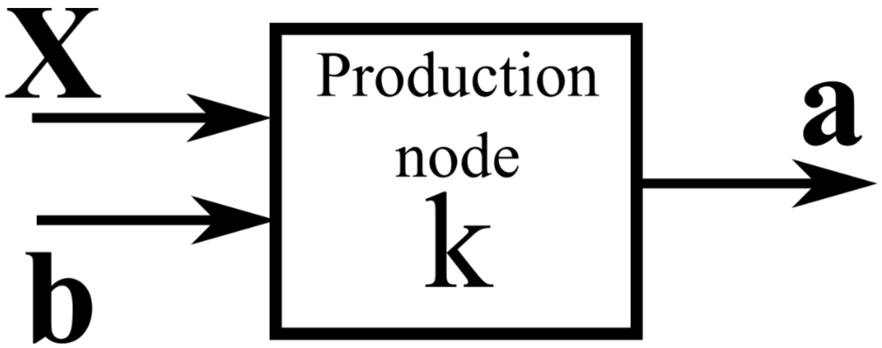
The maximum possible error:
(for small errors only)

$$\Delta X_{\max} = \left|\frac{\partial f}{\partial p}\right|\Delta p + \left|\frac{\partial f}{\partial q}\right|\Delta q + \left|\frac{\partial f}{\partial r}\right|\Delta r + \ldots$$

https://en.wikipedia.org/wiki/Taylor_series

Example 1: Mass balance (addition and subtraction)





Left: a and b are given, X = a + b

$$\Delta X_{\max} = \left|\frac{\partial X}{\partial a}\right| \Delta a + \left|\frac{\partial X}{\partial b}\right| \Delta b$$
$$= |1| \cdot \Delta a + |1| \cdot \Delta b$$
$$= \Delta a + \Delta b$$

Right: a and b are given, X = a - b

$$\Delta X_{\max} = \left|\frac{\partial X}{\partial a}\right| \Delta a + \left|\frac{\partial X}{\partial b}\right| \Delta b$$
$$= |1| \cdot \Delta a + |-1| \cdot \Delta b$$
$$= \Delta a + \Delta b$$

In both cases the maximal errors are the same! ('Worst case': errors add up.)
Left: Both a and b are larger than what was measured. **Right:** a larger, b smaller than measured.

Note: Be careful in situations where X = a – b and a ≈ b. Here, X ≈ 0 but ΔX can be large.

Example 2: Parameter equation with factors

Example: IPAT-equation, linking population, affluence, and technology (all positive) with environmental impacts: $I = P \cdot A \cdot T$

Absolute maximal error:

$$\Delta I_{\max} = \left|\frac{\partial I}{\partial P}\right| \Delta P + \left|\frac{\partial I}{\partial A}\right| \Delta A + \left|\frac{\partial I}{\partial T}\right| \Delta T$$

$$= |AT| \cdot \Delta P + |PT| \cdot \Delta A + |PA| \Delta T$$

$$= \Delta P \cdot A \cdot T + P \cdot \Delta A \cdot T + P \cdot A \cdot \Delta T$$

Relative maximal error:

$$\frac{\Delta I_{\max}}{I} = \left|\frac{\partial I}{\partial P}\right| \frac{\Delta P}{I} + \left|\frac{\partial I}{\partial A}\right| \frac{\Delta A}{I} + \left|\frac{\partial I}{\partial T}\right| \frac{\Delta T}{I}$$

$$= |AT| \cdot \frac{\Delta P}{PAT} + |PT| \cdot \frac{\Delta A}{PAT} + |PA| \frac{\Delta T}{PAT}$$

$$= \frac{\Delta P}{P} + \frac{\Delta A}{A} + \frac{\Delta T}{T}$$

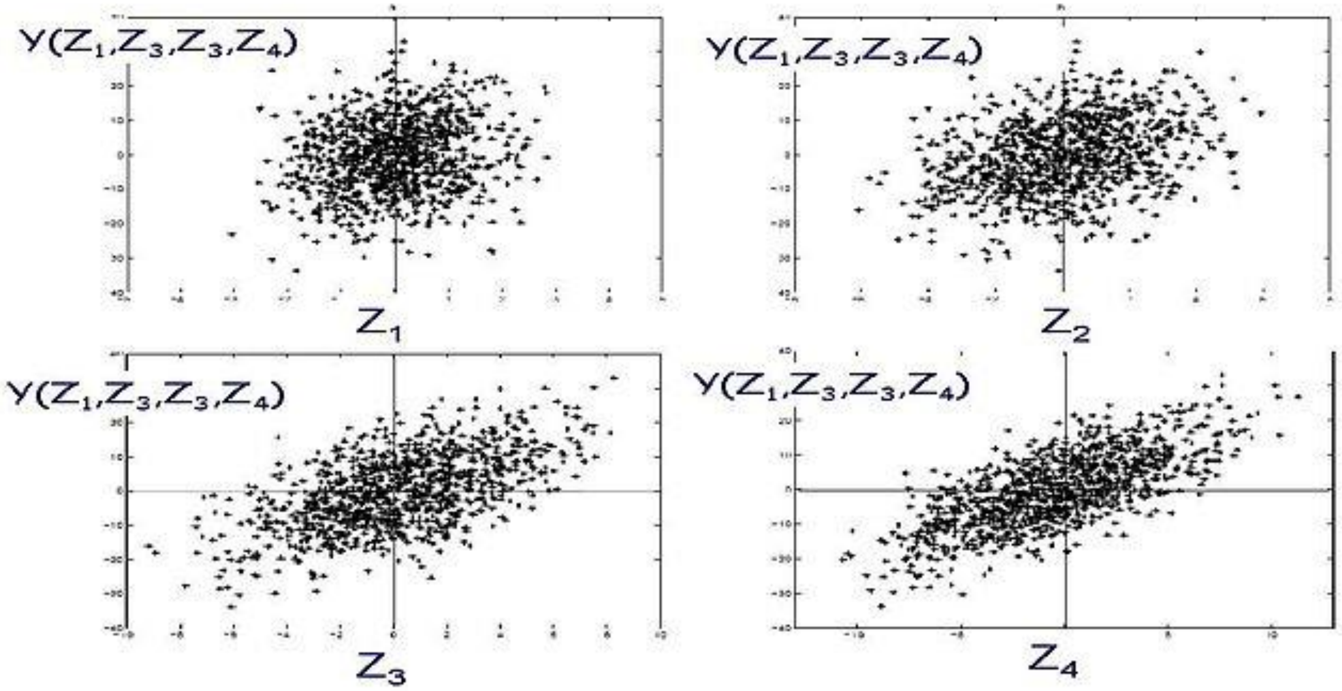
In the case of a parameter equation with factors only, the relative maximal error is simply the sum of the relative maximal errors of the parameters.

If $\frac{\Delta P}{P}, \frac{\Delta A}{A}, \text{and} \frac{\Delta T}{T}$ are 2%, 4%, and 5%, respectively, the resulting $\frac{\Delta \mathrm{I}_{\max}}{I}$ is 11%.

https://en.wikipedia.org/wiki/I_%3D_PAT

d) How do errors propagate from parameters to system variables?

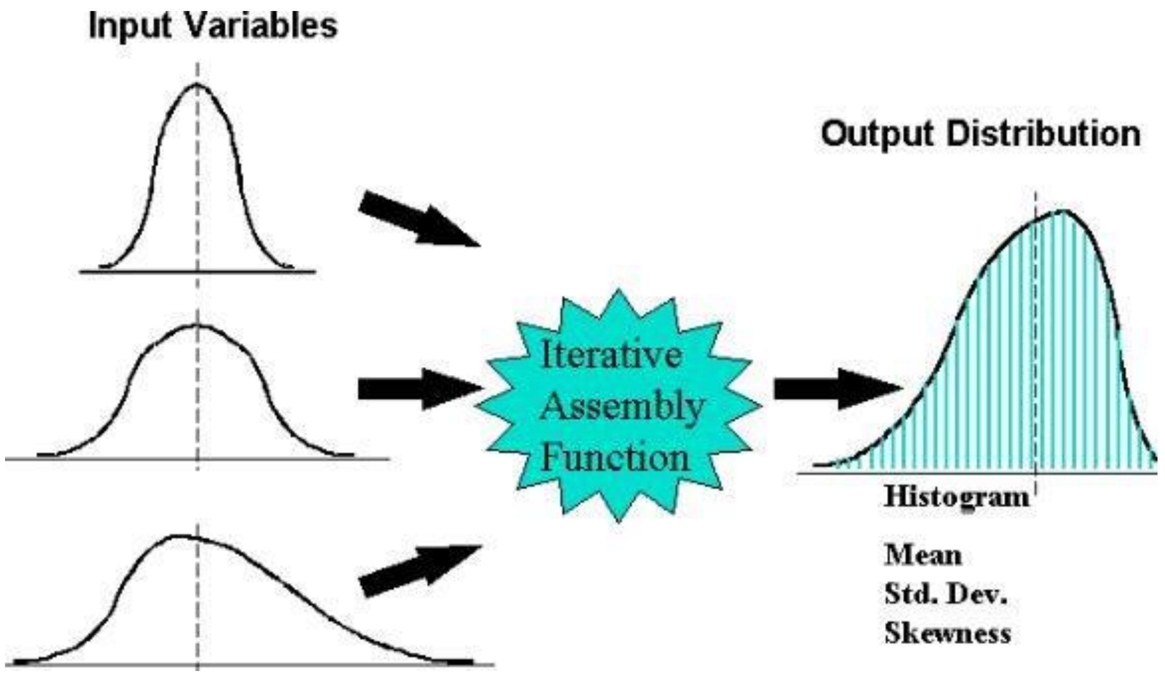
ii) Numerical approach with the Monte Carlo Method



The Monte Carlo Method / *Monte-Carlo-Simulation*

For more complex models and analytical calculation of result uncertainties is not possible or too laborious.
Single-parameter sensitivities may not tell the whole pictures in cases where models are nonlinear.
The solution: The so-called *Monte-Carlo Simulation*:

- 1) Identify the probability distributions of the model input parameters and data.
- 2) Draw a sufficiently large number of random samples for all input variables.
- 3) Compute model output for each sample.
- 4) Treat model outputs as sample and apply statistical analysis to all model results.



https://media-exp2.licdn.com/mpr/mpr/AAEAAQAAAAAAAALqAAAAJDgzN2UxMTU5LWVmNDYtNDQyMC04Y2QyLWZhNDljYjgyZDY4MA.jpg

History of the MC method

- *Buffon's needle experiment:* An early experimental approach to measure π (https://en.wikipedia.org/wiki/Buffon%27s_needle).

- First MC with stochastic sampling carried out in nuclear physics, mainly as part of the different U.S. nuclear weapon's programme, including the Manhattan Project (40ies and 50ies).

- Those activities were all secret, hence a code name was needed to disguise the method!
  → Monte Carlo as name of the Casino de Monte-Carlo in Monaco was chosen.

- Today, it is applied across as wide spectrum of scientific fields to test models and make predictions, f from quantum mechanics to environmental science.

https://en.wikipedia.org/wiki/Monte_Carlo_method#History

When is the Monte Carlo method suitable?

The MC method is suitable for:

- any type and shape of input distribution functions
- any combination of different distribution functions
- large uncertainty ranges
- complex model equations and solutions (LCA!)

The MC method is not suitable for:

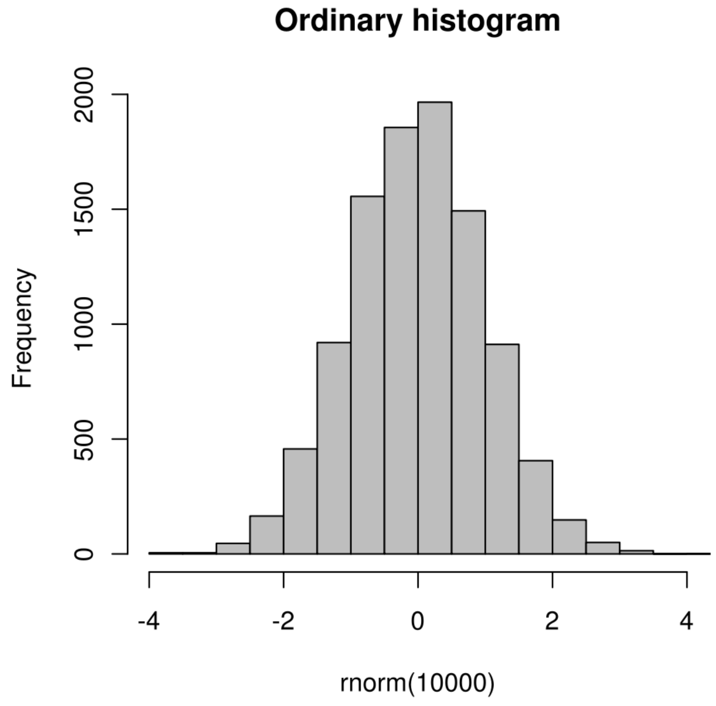
- borderline model cases (slight parameter variation leads to infeasible case)
- Models with very large computation times

Some tips for the MC method I

- EXCEL can be used to perform MC simulations…

  - *"=RAND() / Zufallszahl()"* draws a random number between 0 and 1
  - *"=NORMINV(RAND(), mu, sigma) / =NORM.INV(Zufallszahl(), mu, sigma)"* draws a random number of a normal distribution with mean "mu" and standard deviation "sigma"

- ... and to draw a histogram of the model result



- The mean of the sample is an estimator of the expected value:

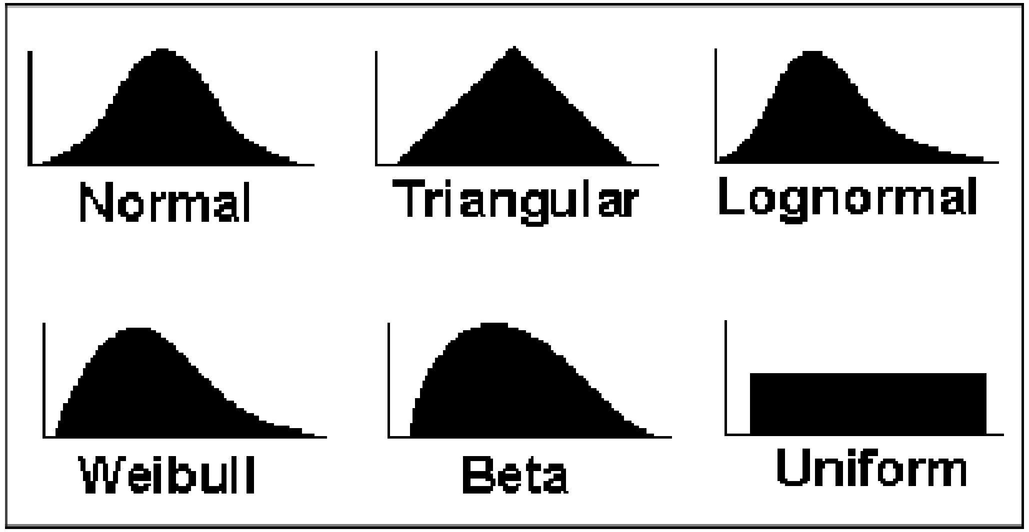
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- The variance (*Varianz*) s^2 or $Var(\{x_i\})$ of the sample is an estimate of the standard deviation *s* of the distribution

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Some tips for the MC method II

Suitable distribution types

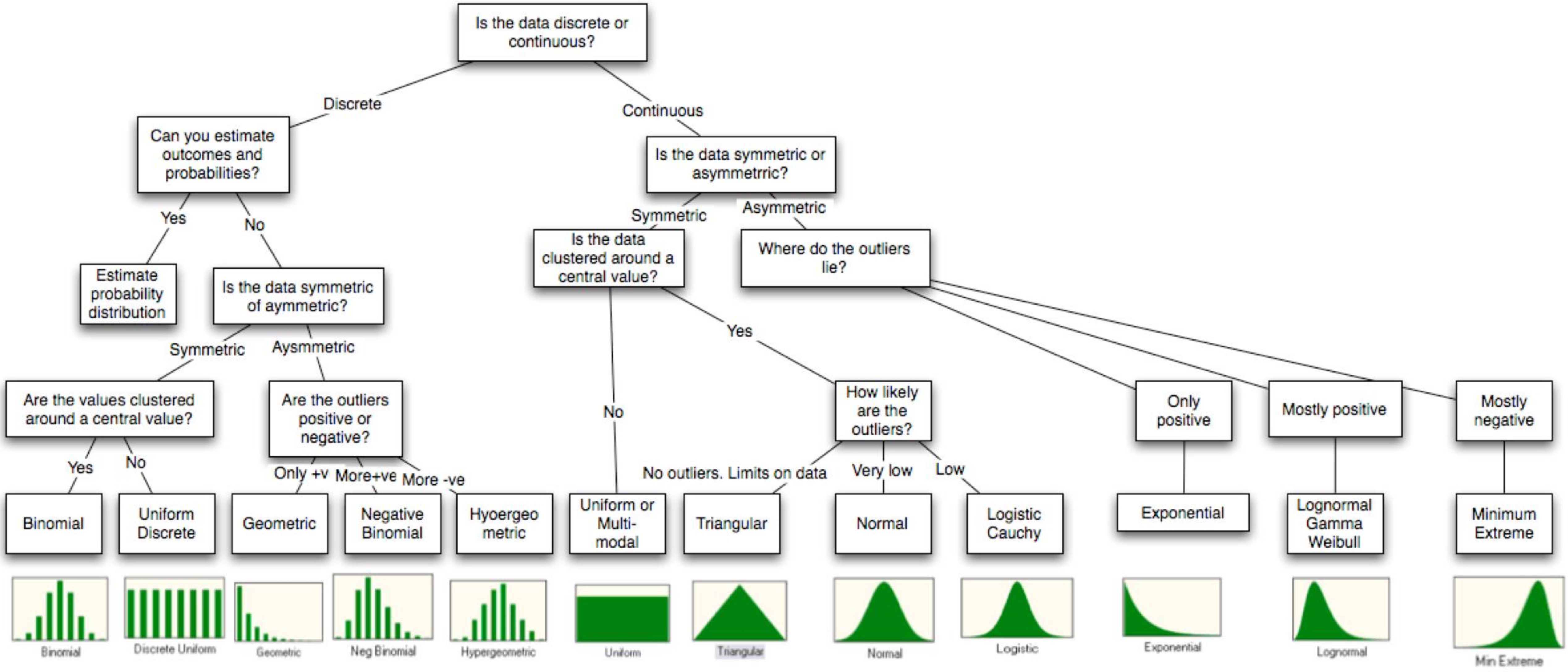


… and many more!

The exact shape of data and parameter distributions is not known in most cases.
Reasonable assumptions have to be made.
“Garbage in – garbage out!”
+ For bounded data a uniform distribution may be suitable.
+ For nonnegative data the lognormal distribution may be suitable.
+ For data with high certainty a normal distribution may be suitable.

https://www.researchgate.net/profile/John_Hayse/publication/228836732/figure/fig1/AS:340598570668044@1458216435758/Figure-2-Types-of-probability-density-functions-Each-probability-density-function.png

Some tips for the MC method III: Distribution choices

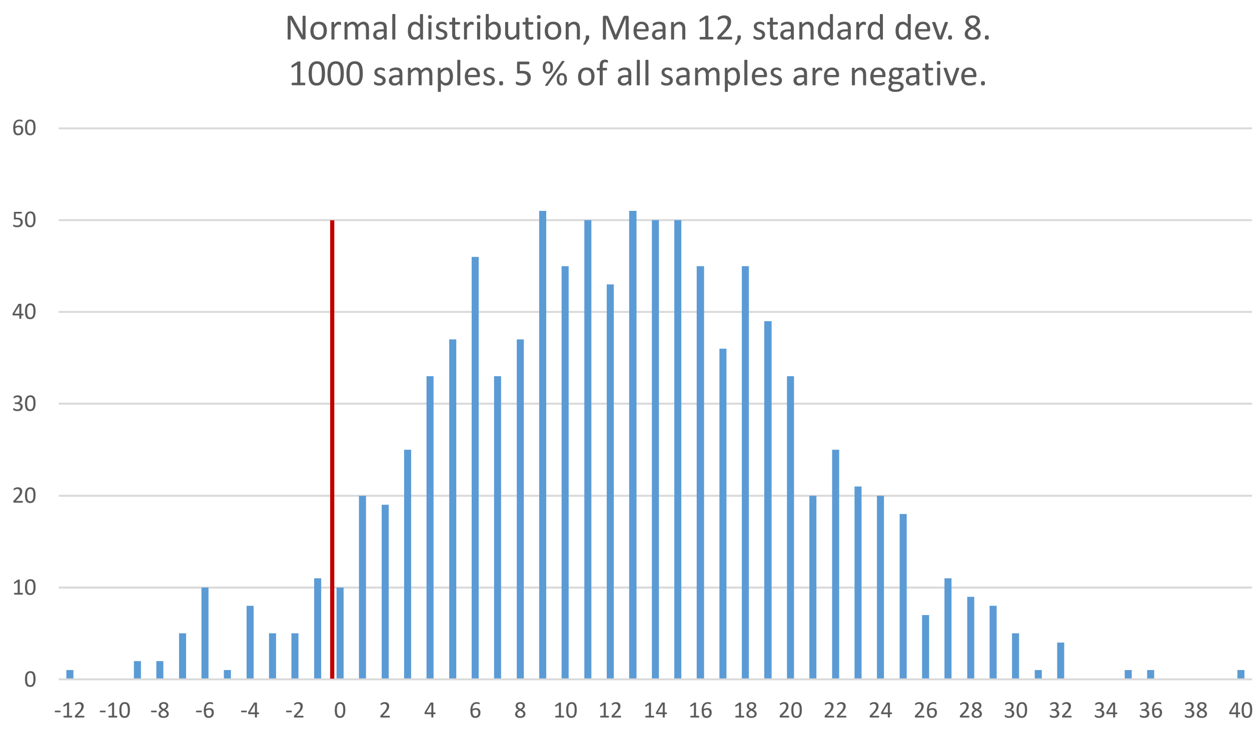


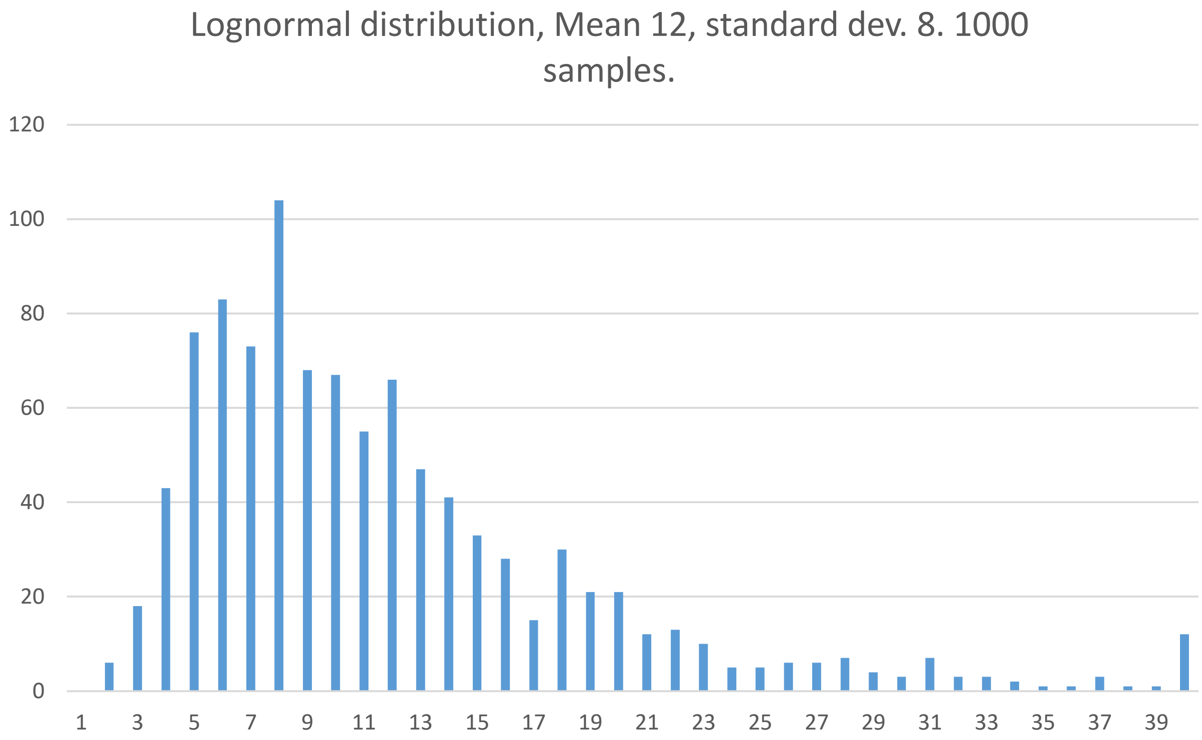
https://www.abstractclasses.in/wp-content/uploads/2017/10/overview-prob-distr.png

Some tips for the MC method IV

Avoiding out-of-scope values: Normal vs. log-normal distribution

Example: normally and lognormally distributed percentage with mean 12% and standard deviation 8%:





Normal distribution:
ca. 5 % of all sample values are negative!

Lognormal distribution:
Long 'tail' of outliers.

3) How to report numeric values and their errors

“We found the copper content of passenger cars to be **26.74383** kg.”

10 milligram precision for copper in an average car. Really?

Some guidelines for how to correctly represent numerical values:

	Plain number	Scientific notation
No error estimate available: Round to 2, max. 3 significant digits or verbal description	435434.334 → 440000 1743 → 1740 or 1700 1746 → 1750 or 1700 198 → ‘about 200,‘ 6.8345 → 7 or 6.8	435434.334 → 4.4E5 1743 → 1.74E3 or 1.7E3 1746 → 1.75E3 or 1.7E3 198 → 2.0E2 6.8345 → 7E0 or 6.8E0
Error estimate available: 1) Round Error estimate 2) Adjust precision of mean so that both are the same	45.345623 ± 0.564 → (45.3 ± 0.6) 123.324 ± 0.14323 → (123.32 ± 0.15) 1743 ± 232 → (1740 ± 240) 1743 ± 560 → (1700 ± 600)	45.345623 ± 0.564 → (4.53E1 ± 6E-1) 123.324 ± 0.14323 → (1.2332E2 ± 1.5E-1) 1743 ± 232 → (1.74E3 ± 2.4E2) 1743 ± 560 → (1.7E3 ± 6E2)

How to correctly round an error estimate?

1) Always round up!
2) If first significant digit is 1 or 2, round to two digits precision. If it is 3 or higher, round to one digit.

Note: In scientific notation, 1.7E3 is not the same as 1.700E3! Trailing and leading zeros are no significant digits!

Overview on model testing (10 different ways!)

	Via inserting alternative values	Via elasticities
Sensitivity analysis	Re-calculate stocks and flows with alternative value, study a small increment, like a 1% increase of a parameter	Linearize the model and calculate the change in the system variable Y due to the change in parameter p via the elasticities dY/dp: $Y_{new} = Y_{old} + \frac{\partial Y}{\partial p} \cdot \Delta p$
Error analysis	Re-calculate stocks and flows not with the central value but with the upper or lower boundary of the error interval.	Linearize the model and calculate the maximal change in the system variables Y due to the change in parameter p via the elasticities dY/dp: $Y_{\max} = Y_{old} + \left\lvert \frac{\partial Y}{\partial p} \right\rvert \cdot \lvert \Delta p \rvert$
Scenario analysis	Re-calculate stocks and flows with alternative value, assuming a substantial change (like use of new technology, major behavioural change)	Does not apply. Large-scale changes are not calculated with locally valid approximations of the model, such as elasticities.

For all approaches: Results can be given in absolute values (changes in stock and flow values in their respective units) or in relative terms (% increase or decrease from the base value of the different stocks and flows).

Take-home messages

- A sensitivity analysis is the least you can do to study and present variability in your model and data! It must accompany every proper analysis.

- Do make an effort, and document it, to understand the nature of uncertainty in both the parameters (error range, distribution?) and the model results (systematic and random errors?)!

- Make use of at least one error propagation technique!

- Numerical values require proper rounding! If no error ranges are present, round to two, max. three significant digits!

Thank you for your attention!

Contact information:

Stefan Pauliuk

Faculty of Environment and Natural Resources

University of Freiburg

Tennenbacher Strasse 4

D-79106 Freiburg, Germany

stefan.Pauliuk@indecol.uni-freiburg.de