UNIVERSITÀ DEGLI STUDI DI PADOVA

DEPARTMENT OF INDUSTRIAL ENGINEERING

Machine Learning
Lesson #16 – Flipped lecture

Academic year 2025-2026

Prof. Pierantonio Facco

CAPE-Lab, Computer-Aided Process Engineering Laboratory

Email: pierantonio.facco@unipd.it

URL: https://research.dii.unipd.it/capelab/

Today's lecture

1/2:

Q&A on the flipped lecture

1/2:

DoE intro

Q&A on the flipped lecture

White slide because now YOU have to build the lecture with your questions!

UNIVERSITÀ DEGLI STUDI DI PADOVA

DEPARTMENT OF INDUSTRIAL ENGINEERING

Design of Experiments Lesson #1

Academic year 2025-2026

Prof. Pierantonio Facco

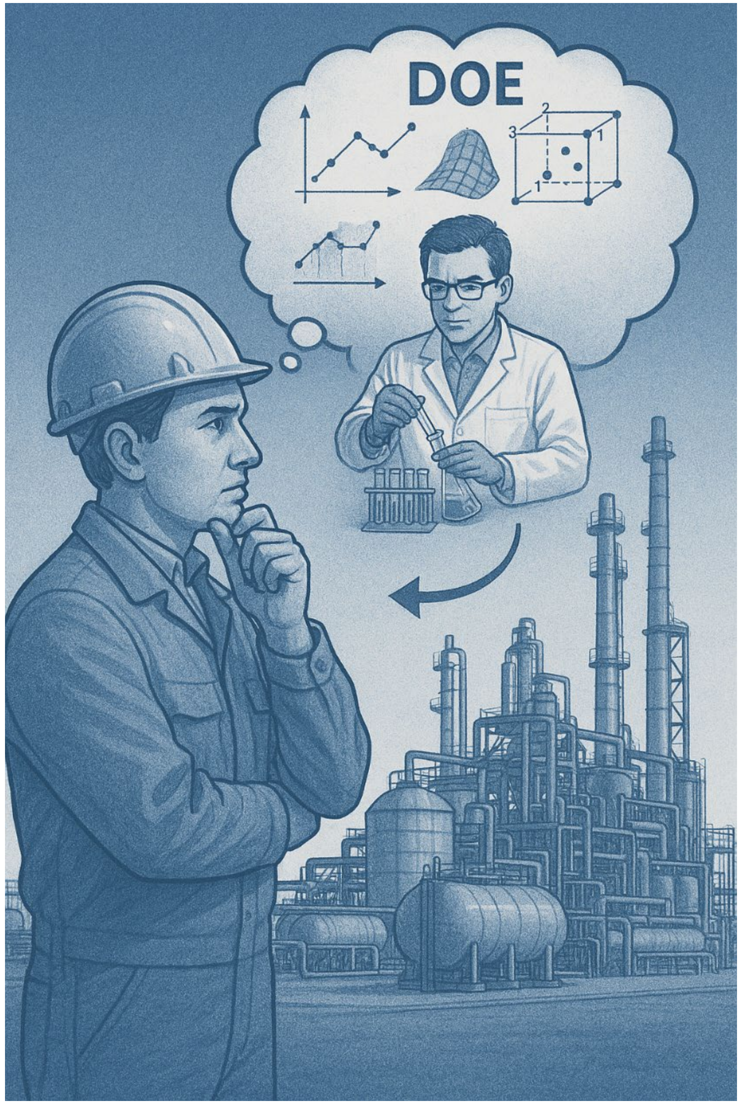
CAPE-Lab, Computer-Aided Process Engineering Laboratory

Email: pierantonio.facco@unipd.it

URL: https://research.dii.unipd.it/capelab/

Objectives of this part of the course

- This course will give you the basis for **science-based decision making in experimental activities** (regardless the experimental work takes place in a laboratory, a pilot plant, or a full-scale production plant):
  - doing **experiments** efficiently
    - define the problem
    - define the experimental domain
    - plan experiments
  - **analyzing data** from the experimentation (having experimental data do not guarantee to know anything about the problem)
    - processing and analyzing experimental data
    - extracting information on the data
  - **interpreting the results** of the analysis
    - building and interpreting a mathematical model describing a process/product
  - translating the interpretation of the outcomes in **action**
    - converting models' interpretations in concrete plans
    - iterating experimentation



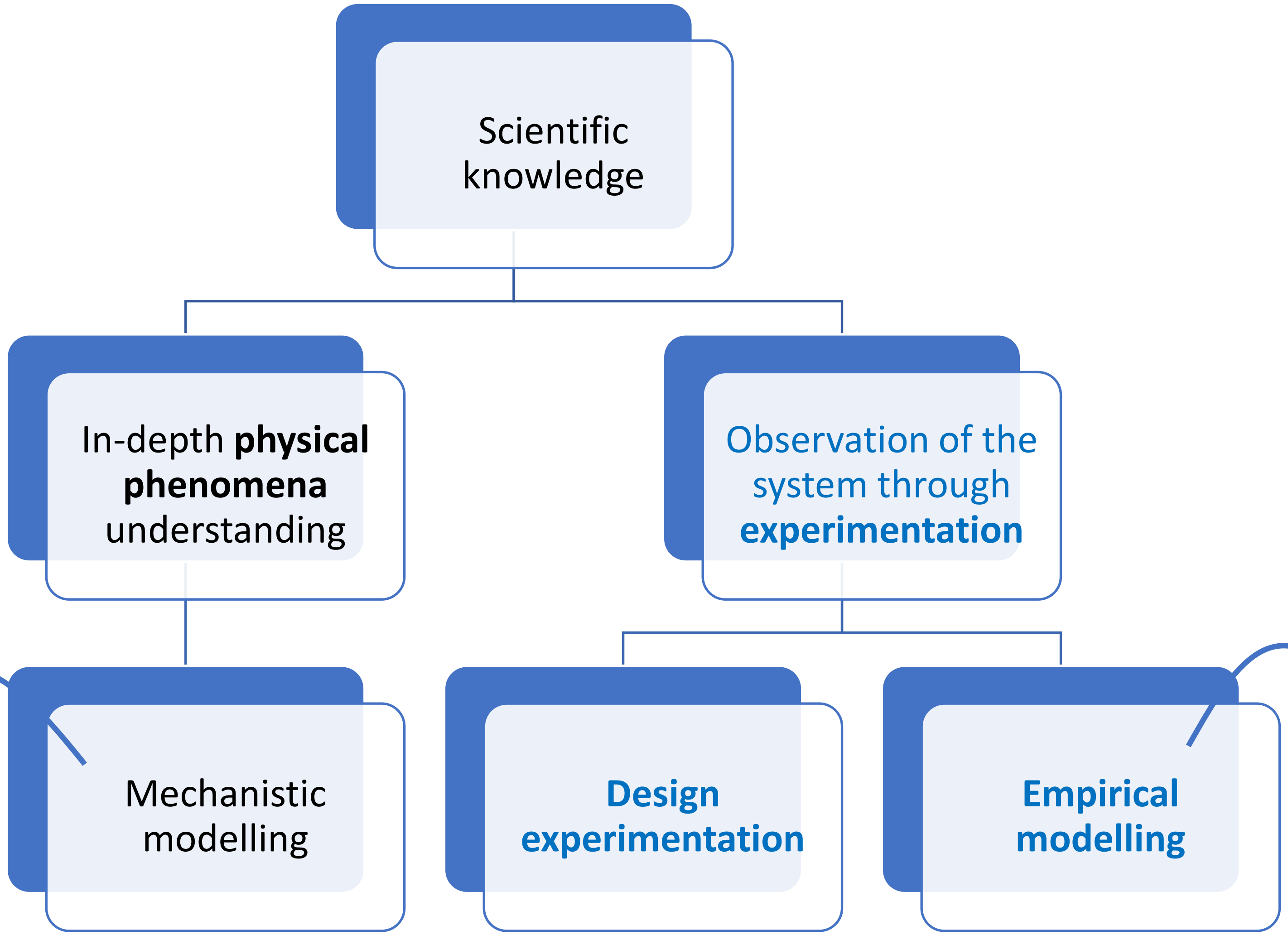
Why experiments?

- Answers to general questions:
  - how can we find optimum of a process?
  - what is the best combination of *factors* (combination of raw materials, inputs, settings, etc…) to ensure a predetermined *response*, i.e., product quality?
- Motivations:
  - development of **new products/processes**
  - **screening important factors**
  - **enhancement** of existing products/processes
  - **optimization**
    - product quality/performance
    - manufacturing procedures
    - analytical instrumentation
  - **minimization** of the effect of external and unknown variability sources, production costs, waste/scraps, reworks, pollution, etc…
  - testing **robustness** of products/processes/analytical instrumentation

Sectors of applications

- Experimentation play a fundamental role in:
  - technology commercialization
  - product realization
    - novel product design/formulation
    - process development
    - process improvement
- Application **sectors**:
  - chemical industry
  - polymer manufacturing
  - plastics and paint industry
  - car production
  - pharmaceutical and biopharmaceutical industry
  - food and dairy industry
  - pulp and paper production
  - steel industry
  - mining industry
  - mechanical industry
  - telecom industry
  - marketing
  - service and business operations
  - etc.

Experimental scientific (engineering) methodology



$$\frac{\partial c_i}{\partial t} = -\nabla(c_i \boldsymbol{v}^*) - \nabla \boldsymbol{J}_i + R_i$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

Experiments

- General objective: understanding the cause-and-effect relationships in a system

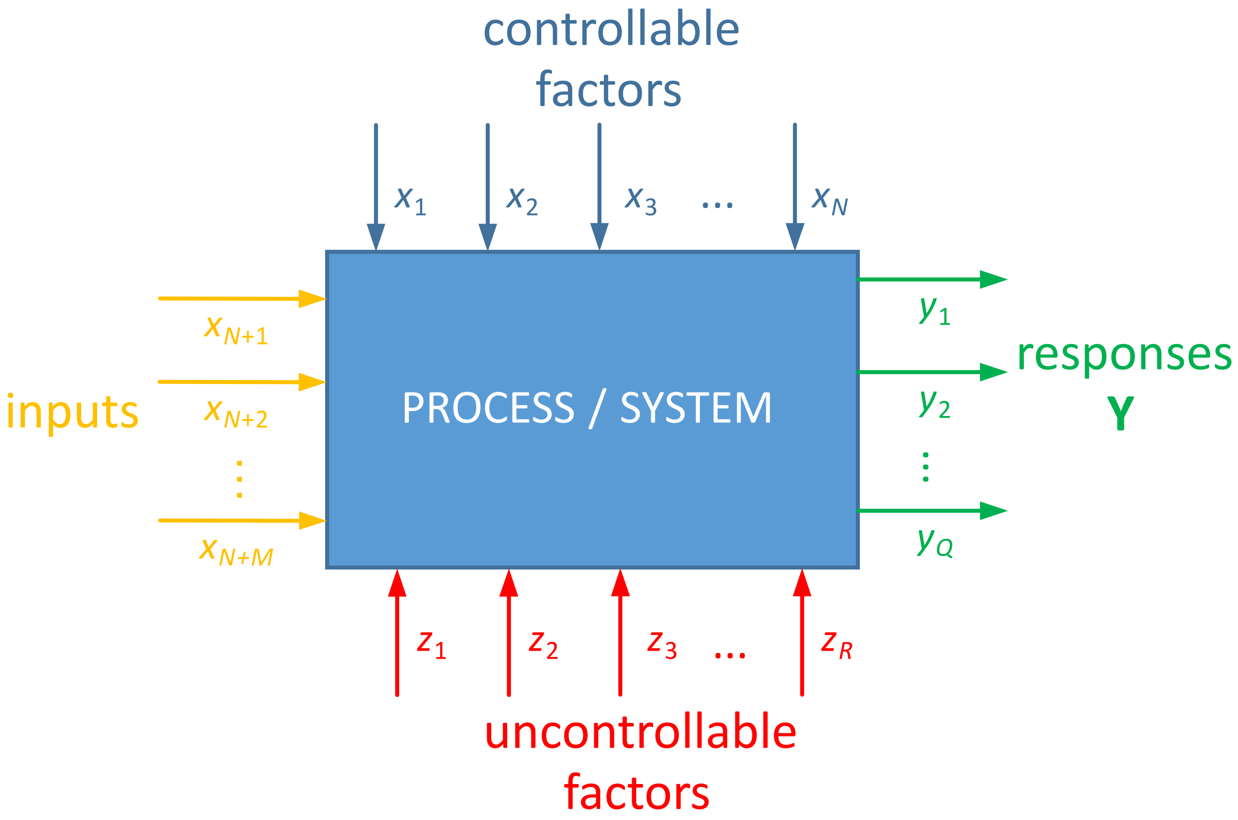
it is required to *deliberately change* the input variables (**factors**) to observe the changes that the inputs determine in the system output variables (**responses**)

Experiment: a test or series of runs in which *purposeful changes* are made to the *input variables* of a process/system in such a way as to identify the reasons for the output *response* changes

- DoE involves making a set of *experiments* which are meaningful with regard to the given question

Processes and experimental strategy

- Experiments are used to study the performance of a process/system
  - understanding the combination of operations, machines, people, resources, settings, etc. that transform the inputs into an output characterized through observable variables
- Variables:
  - inputs
  - controllable variables
  - uncontrollable variables
  - responses



Experimentation strategy

… a little bit of nomenclature

- **Responses**: variables which are highly informative about the properties and the general conditions of the system
  - responses reveal if the process behaves in a proper manner
- **Factors**: variables which manipulate the system under study and exert an influence on the response
  - factors exert an influence on the system that should be mapped

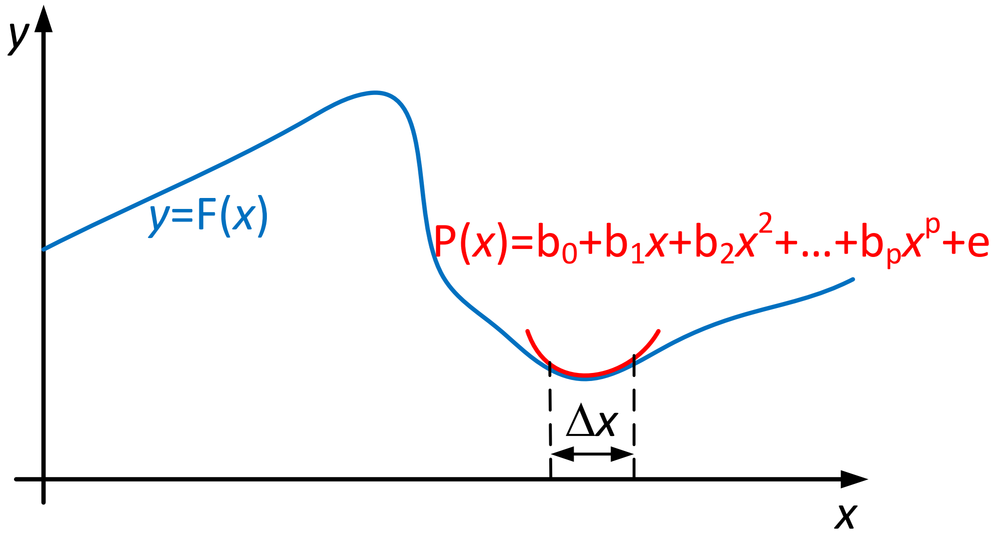
Connect information on factors and responses by means of **mathematical modelling** to understand their **relationship**:

- approximation of the reality which provides a sound tool for understanding important mechanisms of the reality
- allows manipulating parts of the reality according to a desired outcome

Any idea on how to do this?

Theoretical and empirical models

- Modelling strategies:
  - **theoretical modelling**: *hard* models derived from well-established and accepted theories based on the fundamental laws underlying the system
  - **empirical modelling**: *soft* models to describe how factors influence the responses in a predetermined interval
    - valuable alternative when the fundamental mechanism is :
      - not understood very well
      - too complicated to permit an exact first-principles model to be postulated
  - **semi-empirical modelling**: include some prior knowledge on mathematical operations for building a more accurate empirical model



Taylor series expansions are often used in empirical modelling

- a function y can be approximated by a polynomial $P(x)$ in a limited interval (i.e., domain) Δx
- p is the **complexity** of the model:
  - bear in mind **to find the trade-off between model complexity and investigation ranges**

Empirical modelling

- Empirical models are **local models** that describe the situation in detail within the investigated interval, while global models only give a superficial knowledge of large aspects of the reality
  - local models pertaining narrow regions provide in-depth details
  - smooth behaviors can be effectively described (i.e., well approximated) by **simple polynomials**:
    - maximum model complexity $p = 2$ in almost the totality of the cases

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{12} x_1 x_2 + \cdots + \beta_{11} x_1^2 + \cdots + \beta_{22} x_2^2 + \varepsilon$$

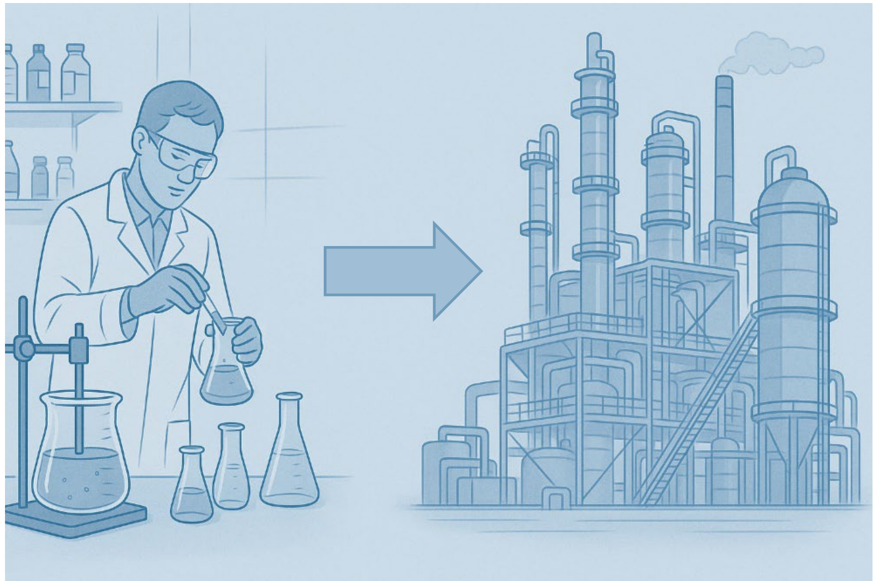
How to conduct experiments?

- Any idea on how to conduct an experiment?
- Let's take the example of a cake…
  - for example, we want to design the most tasteful cake considering the effect of:
    - flour
    - eggs
    - shortening

Ideas?

Example of a chemical reaction

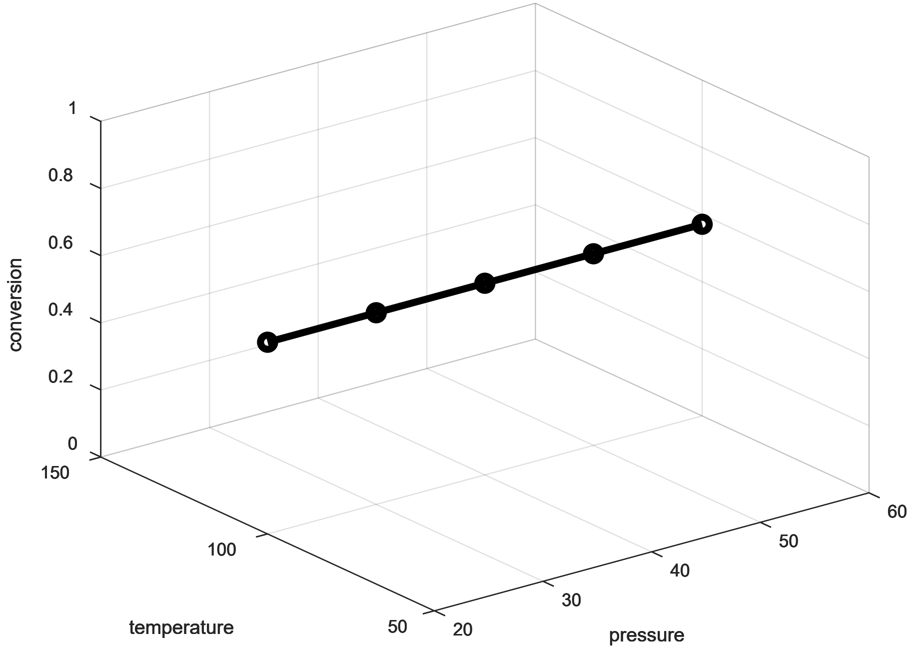
- Consider the case you have to industrialize a product



- The experimentation is carried out changing the 2 most influential factors on the product yield
  - temperature
  - pressure
- Then the domain in which factors are varied is chosen

Example: reaction dependence from T and P (1/3)

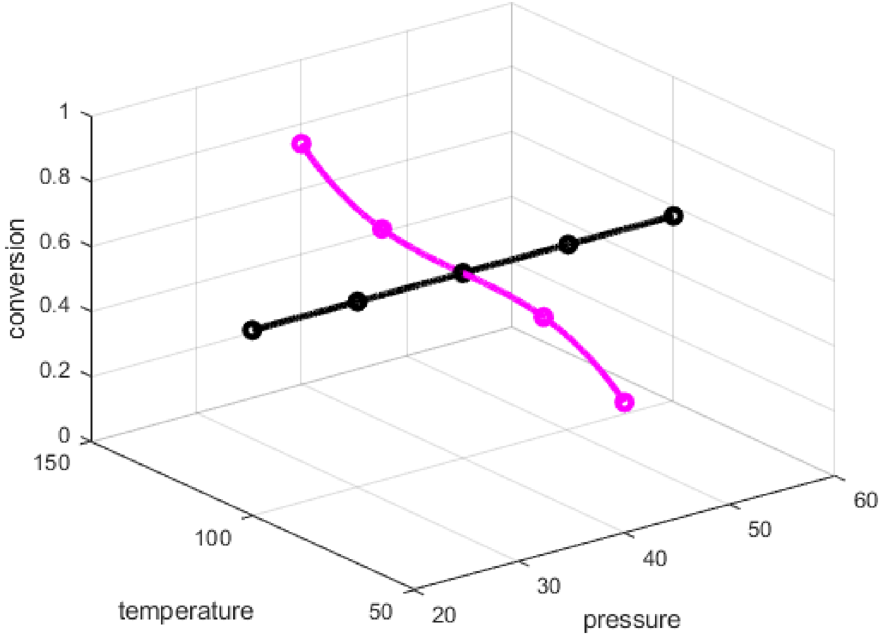
- The reaction is carried out changing:
  - temperature
    - domain: 50°C – 150°C
  - pressure
    - domain: 20 bar – 60 bar
- **OFAT method**:
  - start with pressure at 20 bar
    - keep temperature unchanged at 100°C
  - go up to 30, 40, 50, and 60 bar



- Results:
  - **no sensitivity to the pressure changes!**
  - conversion at 0.57

Example: reaction dependence from T and P (2/3)

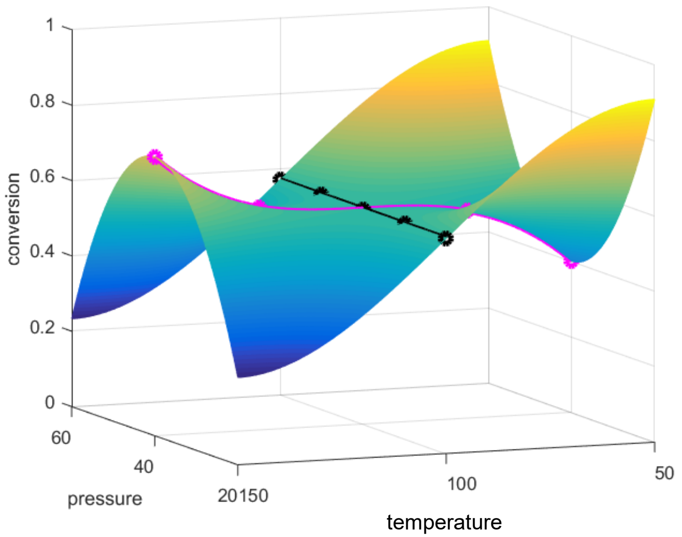
- **Continued OFAT method**:
  - consider the average pressure at 40 bar
  - change the temperature
    - 50, 75, 100, 125 and 150°C

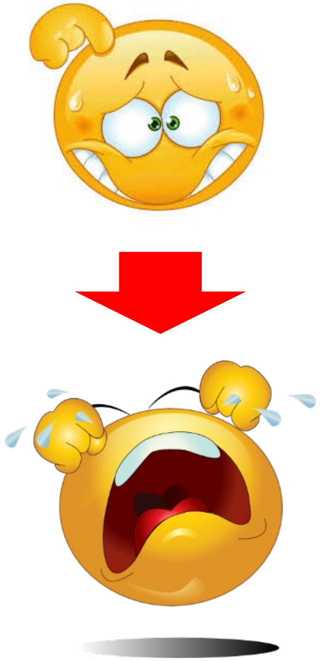


- Result:
  - **sensitivity to temperature**
  - **maximum conversion 0.74**

Example: reaction dependence from T and P (3/3)

- **The sad reality**:
  - the maximum conversion that can be obtained is: 0.95
  - you are losing 28% of the product!
  - you are losing a lot of information
    - pressure influences the conversion
    - there is a strong interaction among pressure and temperature





Change your mind and properly design your experiments!

«Intuitive» approaches

- **Best-guess approach**
  - it is based on previous knowledge and experience of the practitioner

- **One-factor-at-a-time** (OFAT) approach

«Intuitive» approaches: pros and cons

- **Best-guess approach**
  - **pros**: it is easy, since it is based on previous knowledge and experience of the practitioner
    - avoid big deal of technical and theoretical knowledge for experimentation
  - **cons**: the optimal number of experiments can not be determined a priori
    - what if your first guess did not give the desired results?
    - what if your first guess gave an acceptable result? Is it optimal?

- **One-factor-at-a-time** (OFAT) approach
  - **pros**:
    - provides a local knowledge
    - straightforward interpretation of the results
  - **cons**:
    - computationally inefficient distribution of the results
    - poor results
    - fails to consider interactions among factors
    - does not provide a global knowledge, and the content of knowledge is less important than more efficient methods
    - low efficiency:
      - lower content of information with higher number of experiments

Science-based approach to experimentation

- **Statistical Design of Experiments**:
  - **objective approach** to plan the experiments
  - **appropriate data** are *collected* and *analyzed* by statistical methods
  - valid and **objective conclusions**

- **organized approach**: it is a guided procedure for experimentation
- encourages to focus on an experimental objective
- gives the most **parsimonious** experimental plan
- provides **useful and precise information** on the factors' relation and influence on the responses
- can be easily interpreted, also in a graphical manner

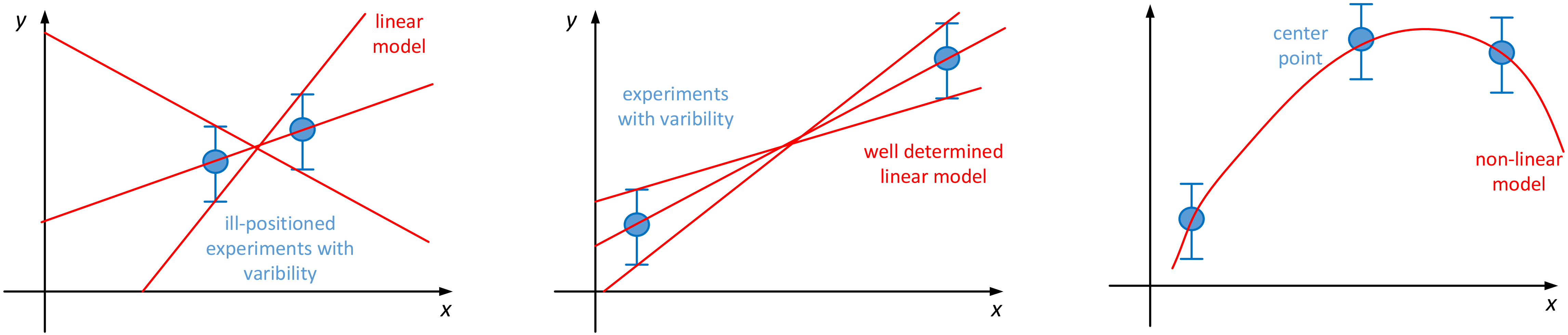
Effectiveness of statistical DoE

- Critical issues faced by statistical DoE:
  - provides informationally **optimal arrangement of the experiments**

  - **higher understanding** on the system
    - **multiple assessment** of the **effect of each factor**
    - estimates correctly the factors' **interactions**
    - evaluates the experimental **domain** along the edges
  - produces reliable and **easy-to-interpret maps** of the system under study
  - considers correctly the **systematic and the non-systematic part of the variability**
    - distinguishes *effects* from *noise*
    - sharpen the estimate of the real effects, thanks to averaging
    - estimates the size of noise through the standard deviation of residuals

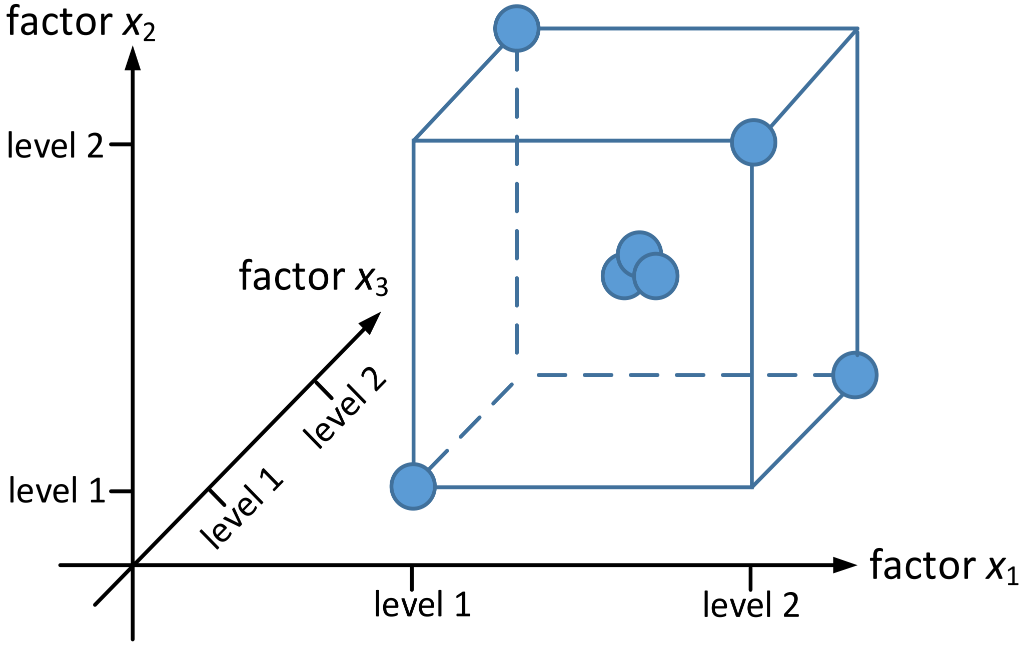
Issues on variability addressed by DoE

- It matters a lot **where and how** the experiments are performed!!!
  - the investigation range of a factor should be considerably larger than the **experimental variability**
    - this guarantees a strong enough «signal» for the factor to be modelled
  - an extra point located in the **domain center** (center point) is favorable to identify non-linearity



Some typical DoE approaches

- **Full-Factorial design**
  - define:
    - the factors which are influential on the response
    - an interesting standard experiment as a center point
    - the levels to be explored and the ranges of the (symmetrical) experimental domain
- 2^K factorial design
- N^K factorial design
- **Fractional-Factorial design**
  - one-half order fractions
  - higher order fractions
- Composite design
- Mixture design
- Etc…



DoE tasks

EARLY STAGES

1. **Discovery**
   - determine what happens when exploring new: materials, factors, ranges
2. **Screening** (not complicated, extremely important when a process/technology is new, usually requires few experiments)
   - uncover the most influential *factors* affecting the process/product
   - determine in which *ranges* the factors should be investigated

CONSOLIDATION

3. **Optimization** (complex task, requires much effort)
   - determine a model to predict the response from all the possible combinations of factors
   - define which combination of the most important factors will result in optimal products/operating conditions
   - finding a possible compromise among (partially) conflicting requirements on the *responses*
4. **Confirmation** (could be extremely useful in scale-up)
   - verify that the system operates/behaves in a consistent way with respect to theories/past experience
5. **Robustness testing** (last tests before the product/method release)
   - determine the sensitivity of a product/process to small changes in the factors' settings (fluctuation occurring in a «bad day»)
   - ascertain that the process/system is robust to small fluctuations
   - ensure robustness to the product/process

COMPLEXITY

DoE stages

- Different objectives for DoE are related to different stages of the experimental process:
  1. **familiarization**:
     - for entirely new applications/equipment
     - spending maximum 10% of the resources (money, time, dedicated personnel, equipment, materials, etc…)
     - usually performed with (simple) full-factorials on a couple of predominant factors with center points
     - verify if replicates give the same results, both in the center and in the corners of the experimental domain
     - useful to develop the measurement technique for the response
  2. **screening**:
     - identify the dominating factors among many factors and their optimal ranges (80/20 rule of the Pareto principle)
     - fractional factorial designs are typically used
  3. **finding the optimal region**:
     - decide where to move the experimental domain to include the optimum
     - use the polynomial models of screening designs (steepest ascent)
  4. **optimization**:
     - many experiments are performed on few selected factors
     - composite designs are usually utilized
     - response surface models are proper tools to:
       - predict the response for a determined combination of factors
       - identify the optimum
  5. **robustness testing**:
     - carried out before the release of the finished product as a last test to ensure quality
     - minimize the factors that in principle have little effect (e.g.: environmental conditions, raw material variability, etc…), but may cause undesired spread around the ideal result through fractional factorial designs
  6. **mechanistic modelling**:
     - utilized to establish a theoretical model, but less utilized in industry

Basic principles

- **Randomization**
  - **both the allocation of the experimental material and the order in which the individual runs of the experiment are to be performed are randomly determined**
    - statistical methods require that the observations (or errors) are independently distributed random variables
      - randomization usually makes this assumption valid
    - “averaging out” the effects of extraneous factors
      - removing systematic bias
    - complete randomization could be difficult, expensive and time consuming
- **Replication**
  - making **independent repeated runs of each factor combination**
    - allows determining experimental error
    - allows obtaining more precision
  - avoid holding constant factors levels in consecutive experiments
  - replication is not measurement repetition
    - **repeated measurements** reflect the inherent variability of the measurement system or gauge
    - **replications** reflect variability sources both within and between runs
- **Blocking**
  - separate experimental runs based on levels of nuisance factors
    - reduces or eliminates variability from nuisance factors
    - improves the precision of comparing factors
  - one block is a set of relatively homogeneous experimental conditions

Take-home message

- Performing experimental campaigns is vital to discover new processes/products, understanding how factors determine the variability of a response, what are the interactions among factors, what are the optimal conditions to run a process or to obtain a desired product
- **DoE** suggests how to perform **optimal experimental campaigns**:
  - parsimonious number of experiments
  - maximally informative experiments that explore well the experimental domain
- DOE allows building **response surface models** (empirical, regression models) that represent the cause-and effect relations between inputs and outputs:
  - in a simple manner
  - with nice graphical interpretation
- We understood the importance of:
  - having a sufficiently **large design space**
  - exploring the **borders** and the **center** of the experimental domain (i.e., design space)
  - **replicating**, **randomizing** and blocking experiments

… per sempre a fianco a me!