Università degli Studi di Padova
Department of Industrial Engineering

Design of Experiments Lesson #2

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/

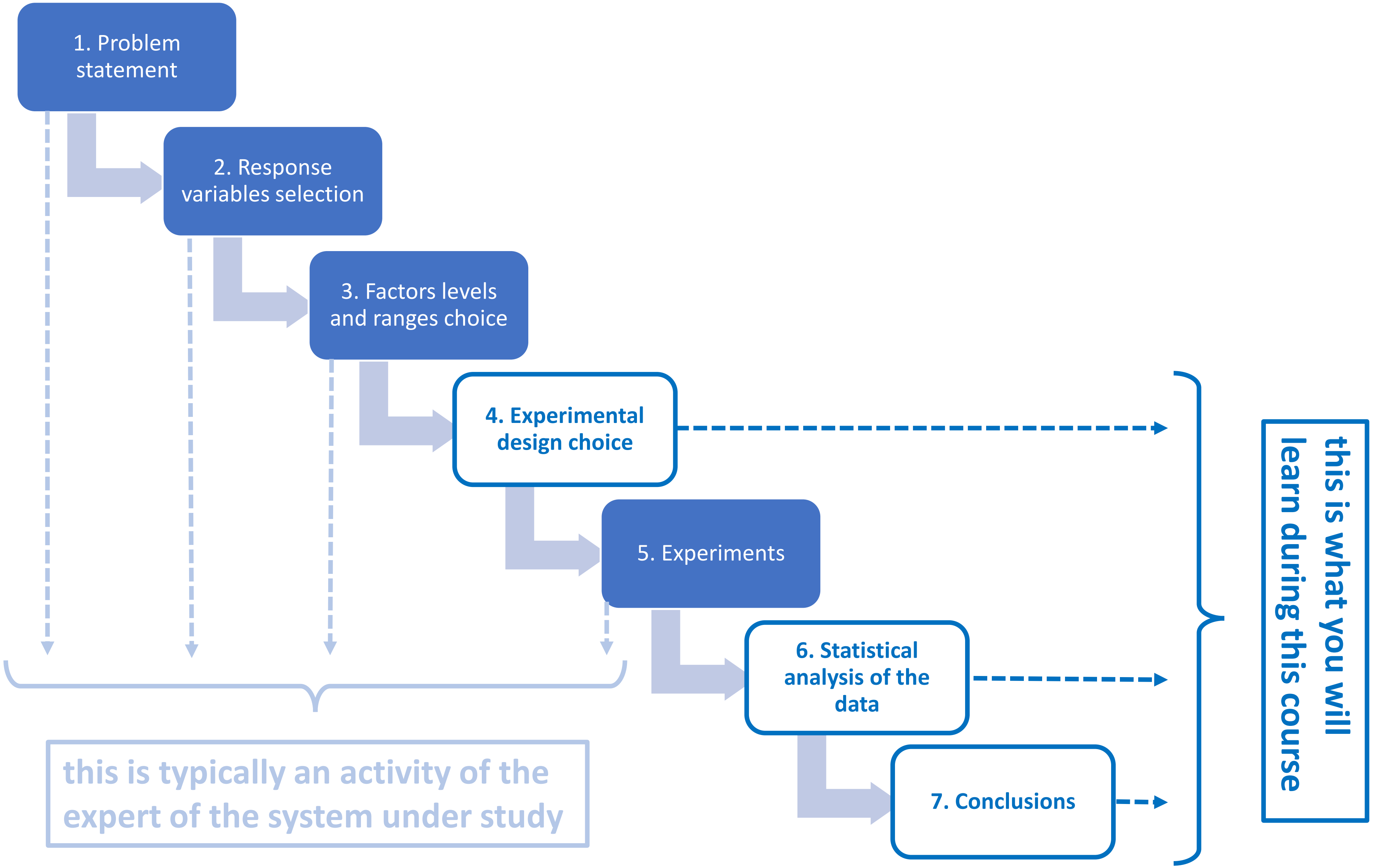
Recap of the last lesson

- A clear experimental **objective** is needed to identify of the relation among a response and (hopefully all) the factors that influence it
- Developing an **informative experimental strategy**:
  - identification of the responses
  - identification of the factors and the experimental domain
  - choice of the optimal compromise among:
    - available resources
    - desired information from the experiments
- Perform the designed **experimental campaign**
  - **randomization**
  - **replication**
  - **blocking**
- **Analyzing data** through **empirical modelling**
- **Implementing solutions**

Today’s lesson

- Science-based procedure to design, carry out, model and analyze experiments
  - choice of responses, factors and experimental domain
  - choice of the experimental design
  - experimental campaign and data analysis through response surface models

Procedure for designing experiments



1. Recognition and statement of the problem

- Fix the overall and the **specific objectives** of the experimentation
  - **team approach** to design experiments: solicit inputs from all the involved parties
    - management
    - marketing
    - customers
    - quality assurance
    - engineering
    - manufacturing
    - operating personnel
- Prepare a list of the **specific problems**
- **Sequential approach**:
  - series of smaller experimental campaigns with specific objectives

2. Selection of the response variables

- **Response variables** (qualitative/quantitative) provide valuable information about the process under study
  - average and standard deviation of the response could be good
  - multiple responses are often necessary
  - categories:
    - regular response: standard variable that can be measured
    - derived response: artificial response calculated as a function of other measurements
- Determine **how the response variables should be measured**
  - consider carefully the calibration and the maintenance status of the measurement system
- The **measurement error** is an important factor
  - if gauge capability is poor consider repeating several times the measurements
- Decisions on response variables should be taken *before* starting the experiments

3.1 Choice of factors

- Identify all the potential design factors
  - factors that the experimenter/process expert may wish to vary during the experiments
  - classify factors in:
    - **design factors**: selected for the study
    - **nuisance factors**:
      - large effects on the responses
      - they must be accounted for
      - classification:
        - **controllable**: its levels can be set by the experimenter → blocking can deal with it
          - different raw materials, different days of the week, etc.
        - **uncontrollable**: cannot be manipulated, but can be measured → analysis of covariance is used to compensate this effect
          - environmental humidity
        - **noise**: natural and uncontrollable fluctuation that is not systematic → robustness studies usually minimize the noise effects
    - **held-constant factors**: despite exerting some effects on the response, they are not interesting for the purpose of the experimentation
    - **allowed-to-vary factors**: factors that are applied in a nonhomogeneous fashion (e.g.: differences among operating units, effects of materials, etc…)

effect on the response is considered to be relatively small

Other categorizations of the factors

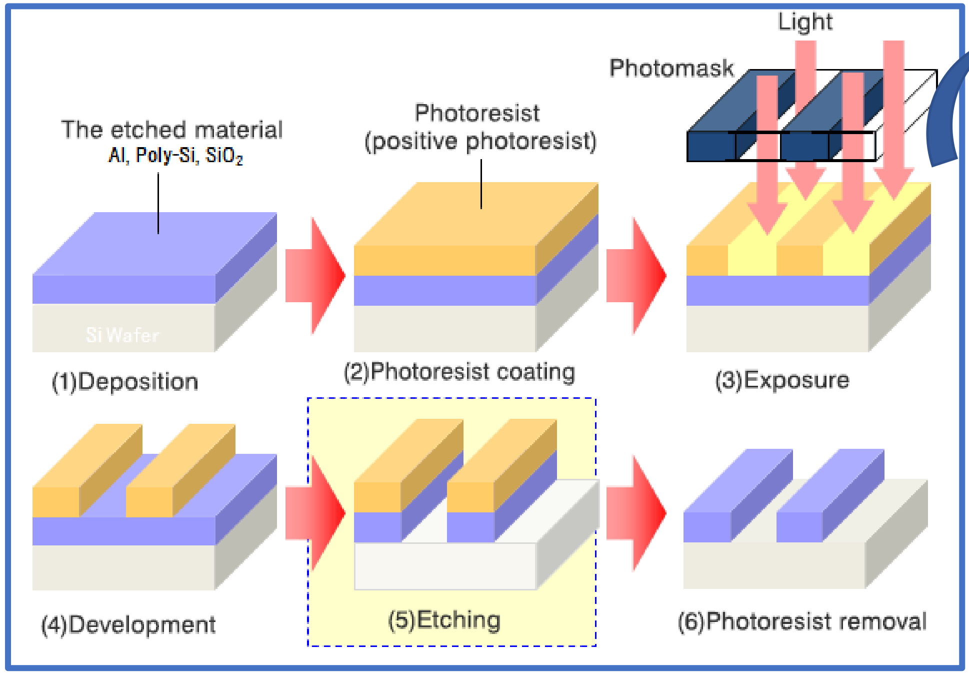
- Process and mixture factors:
  - **process factors**: factors that are manipulated independently of one another
    - power of the laser welding, air mass in engine combustion, pH in chromatography, etc.
  - **mixture factors**: factors which display the amounts of a mixture constituents and add to 100%
    - ingredients in food (e.g.: bakery products, cakes, etc…), concentrations in fine chemicals, etc.
    - *mixture design* is a discipline of the DoE
- Qualitative and quantitative factors:
  - **quantitative factors**: factors that change according to a continuous scale
  - **qualitative factor**: a categorical variable, which can only assume certain discrete values
    - kind of flour used in making a cake, suppliers of raw materials, etc.

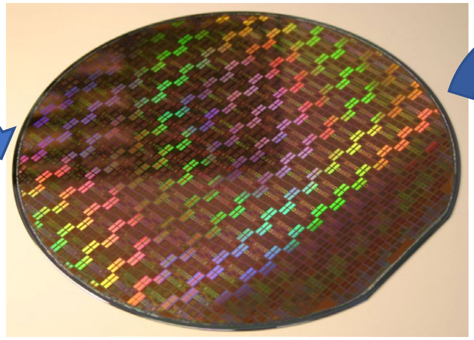
3.2 Choice of levels and ranges

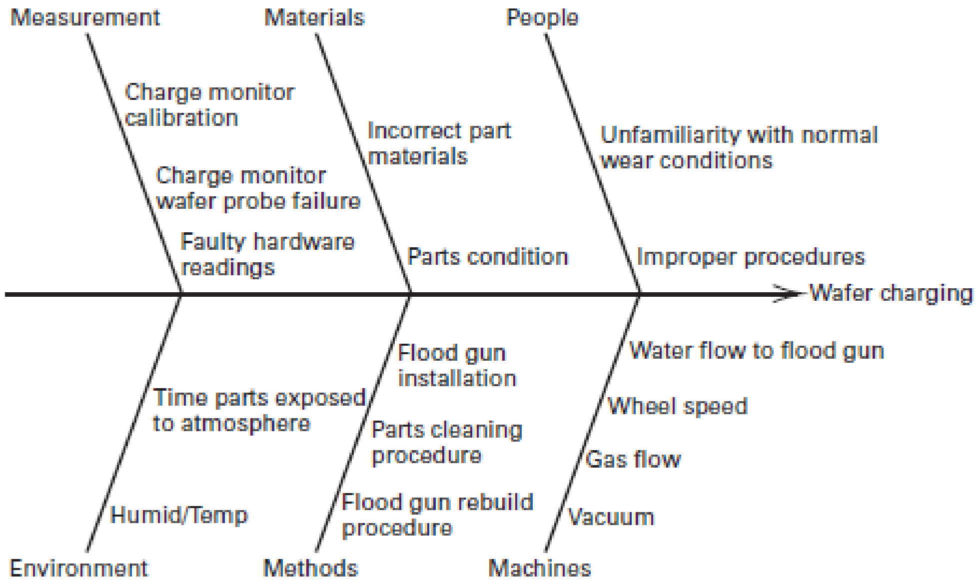
- Choose the **experimental domain** and the region of interest for each variable and keep it sufficiently large
  - use the process knowledge
    - practical experience, **not being overly influenced by past experience**
    - theoretical understanding
  - consider:
    - ensuring experimental feasibility
    - pursuing experimental objective (usually, screening designs require larger domains)
    - evaluating experimental noise
- Choose the **ranges** over which the factors are varied and the **specific levels** at which the experiments are run
  - select the way of *controlling* that the factors are maintained at the desired levels
  - choose the measurement system for the factors
  - the number of factors levels is usually kept **low**
  - consider factor transformation (e.g.: logarithmic transformation, square root, etc…) to linearize the problem
- Use **cause and effect diagrams** to organize some pre-experimental planning
  - fishbone diagrams to understand the effects of:
    - measurements, materials, people, environment, methods, machines, etc…

Example of a cause-and-effect diagram in etching

- Etching process for the manufacturing of semiconductors in the production of wafers of integrated circuits

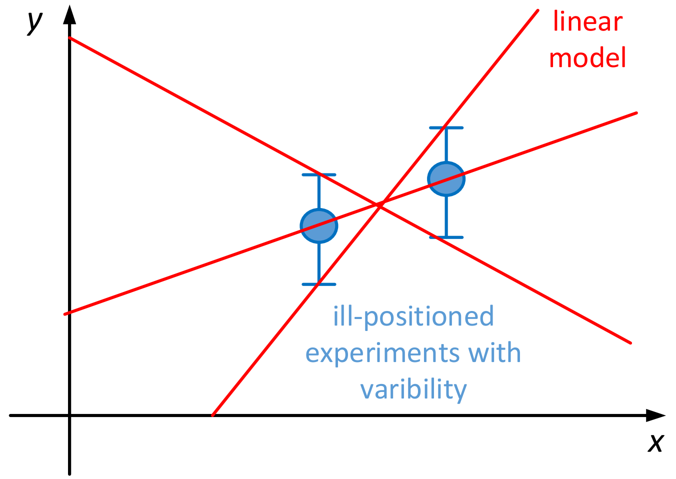


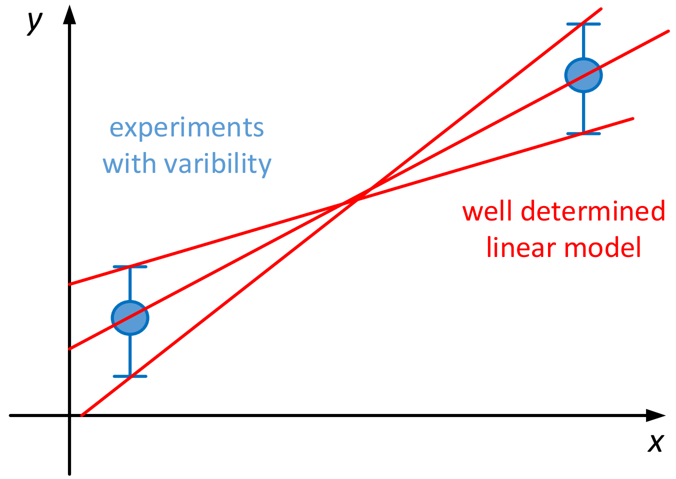


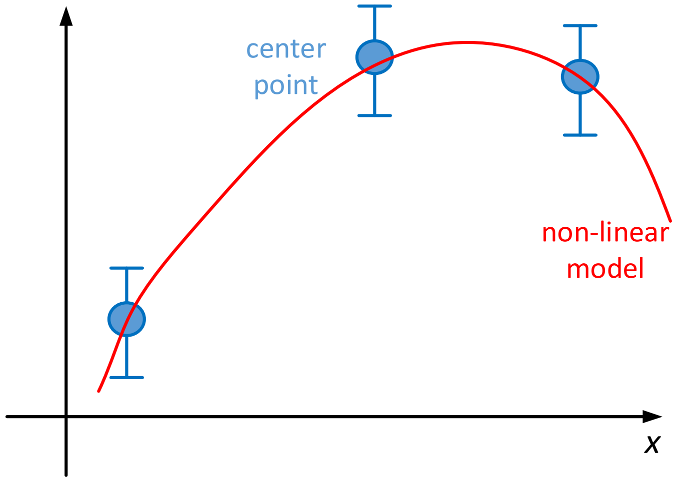


Issues on variability addressed by DoE

- 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» tor the factor to be modelled
  - an extra point located in the **domain center** (center point) is favorable to identify non-linearity







4. Choice of the experimental design (1/3)

- Find the optimal compromise between competing needs:
  - decide the **resources** that could be allocated for the experimental campaign:
    - **time** required for one experiment compared to the time available to obtain results
    - **cost** of the experiments (e.g.: raw materials, dedicated personnel, etc.)
    - number of experiment **replicates** needed
    - **order** of the experimental trials
      - managing blocking and randomization
  - think about and select a tentative **empirical model** for describing the outcomes
    - *statistical packages* aid with modelling:
      - proprietary: JMP, SPSS, SAS, Stata, etc.
      - open source: Python, R, GNU Octave, etc.
      - selected for the course:
        - **Matlab**®, Mathworks Inc. (U.S.A.)
        - **Minitab®**

Do not trust blindly any software you will utilize!!!

4. Choice of the experimental design (2/3)

- **Empirical model**: equation(s) describing the quantitative relation among factors and responses
  - **first-order polynomials** account for the main effects (simple model used in **screening and characterization**):
    - x's are design factors
    - y's are responses
    - β_i are parameters to be estimated

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

  - first-order models with **interactions** (widely used for **in-depth screening**):
    - cross-product terms identify interactions
    - higher order interactions can be included

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

WARNING: the higher the number of model parameters is, the higher the number of the needed experiments

  - **second-order models** adequate for **optimization**:
    - requires more experiments

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

  - **higher order models are rarely necessary**
    - could be suitable for *biological systems* and living systems (e.g.: algae cultivations)

27/04/2026

4. Choice of the experimental design (3/3)

- **Design generation**:
  - **design and chosen model are intimately linked**
  - may be done by means of a commonly available commercial software
- Creation of an **experimental worksheet**:
  - reports the **selected experimental design**
  - additional information:
    - **randomized order** in which to perform the experiments
    - name the experiment
    - annotate also uncontrollable factors

We will learn how to do this in few lessons!

How to select the right design strategy?

- Find the **compromise** among contrasting needs:
  - the **experimental burden**
    - time
    - cost
  - the **objective of the experimental campaign**
    - screening designs usually need fewer experiments to identify the **main effects of the factors** that are most impacting on the response

$$y = \beta_o + \sum_{i=1}^{I} \beta_i x_i + \varepsilon$$

    - assessment of information on how factors **interaction** impact on the response

$$y = \beta_o + \sum_{i=1}^{I} \beta_i x_i + \sum_{i=1}^{I} \sum_{j=i+1}^{I} \beta_{ij} x_i x_j + \varepsilon$$

    - optimizing the system under study needs a wider experimental effort to identify the **high order terms**

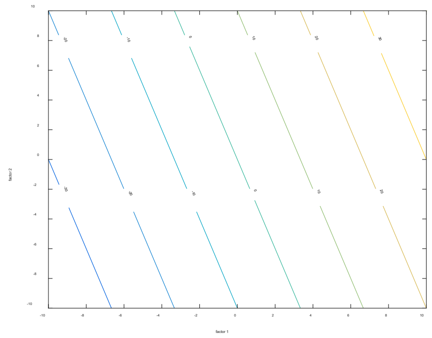
$$y = \beta_o + \sum_{i=1}^{I} \beta_i x_i + \sum_{i=1}^{I} \sum_{j=i+1}^{I} \beta_{ij} x_i x_j + \sum_{i=1}^{I} \beta_{ii} x_i^2 + \cdots + \varepsilon$$

- Stated in mathematical terms, the higher the *complexity* of the response surface, the larger the *number of the model parameters* which need to be estimated, the larger the *number of needed experiments*

Response surface models

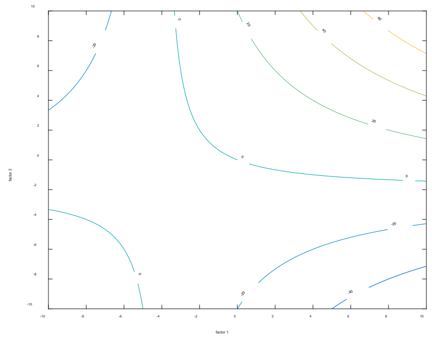
- Linear model

screening:
- what the important factors are
- where the domain of valuable y is



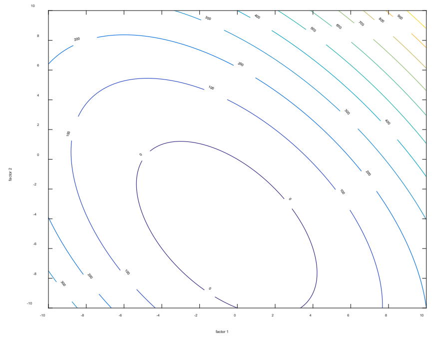
- Linear model with interactions

in-depth understanding:
- what are the interactions between factors that influence y



- Quadratic model

optimizing:
- where are the conditions which maximize/minimize y



5. Carrying out the experiments

- Conduct, if possible, preliminary **pilot runs** which are helpful to:
  - check *measurement system*
  - verify consistency of the materials
  - *evaluate the experimental error*
  - reconsider the *experimental methods and techniques*
- Verify and monitor the process carefully to guarantee that the **experiments are performed exactly according to the plan**:
  - up-front planning is crucial to:
    - prevent mistakes
    - ensure the success of the experimental campaign
- **Perform experiments**!

6. Statistical analysis of the data

- Analyze data obtained from the experiments by means of **mathematical and statistical methods**
  - **graphical methods** help very much to visualize and interpret the outcomes of the study
  - evaluation of the empirical model:
    - analyze in a systematic manner the relation among factors and responses
    - residual analysis and model adequacy help to judge the model
  - the resulting analysis will lead to **objective conclusions** for **decision-making**:
    - measure the *likelihood* of the conclusions and the confidence of the outcomes
    - coupling the results with *engineering judgement and process knowledge* lead to sound conclusions

Today homework

- <u>Exercise #1</u>:
  - try to familiarize with the shape of linear, interaction and quadratic models using Matlab®
    - build a response surface for a linear model
    - build a response surface for a linear model with interaction
    - build a response surface for a quadratic model
    - try the differences in different domains
    - hints: for visualization use the following commands:
      - `contour`
      - `surf`
      - `mesh`

- <u>Exercise #2</u>:
  - search for Matlab® commands that are related to DoE full-factorial and fractional-factorial designs

7. Conclusions

- Draw **practical conclusions** from the performed experiments and the statistical analysis of the data
  - a course of action is always recommended
  - results should be validated through follow-up runs and **confirmation tests**
- Experimental campaign should be **iterative**:
  - avoid large and comprehensive experiments
  - as experimental plan progresses:
    - some variables are dropped
    - some new variables are considered
    - the experimental domain of the factors are varied
    - new response variables may be added, etc.

⇩

- **sequential experimental campaigns** are preferable ($\sim 25\%$ of the resources per iteration)

Sequential experimental campaigns

1st EXPERIMENTAL STEP: screening design

- selection of the important factors
- choice of the proper domain

2nd EXPERIMENTAL STEP: in-depth understanding

- understanding what are the interactions between factors impacting y

3rd EXPERIMENTAL STEP: system optimization

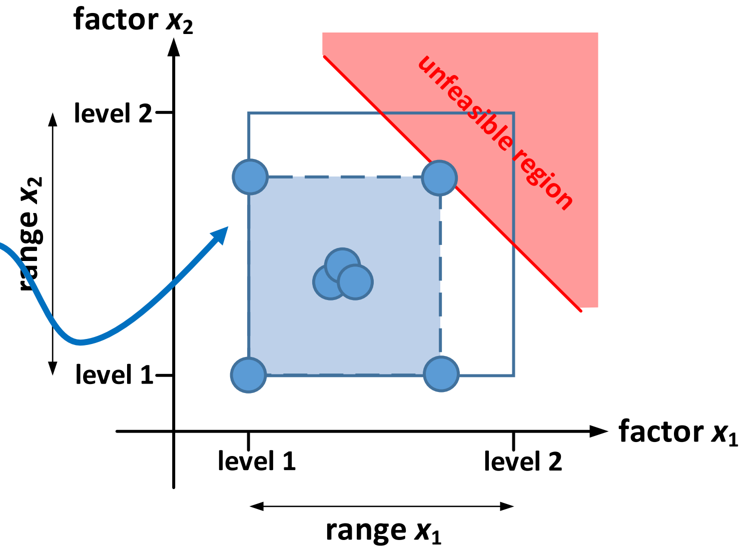
- finding the conditions which maximize/minimize y

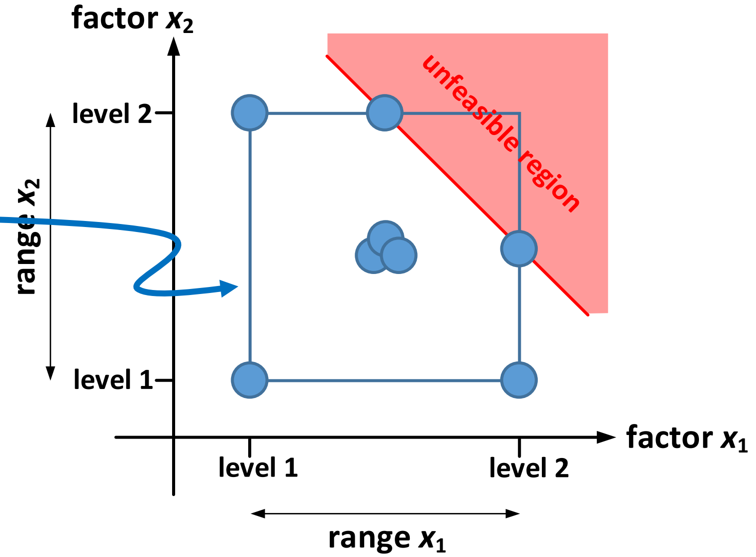
Suggested experimental plans

- Planning a properly designed experimental campaign requires:
  - **fractional factorial design** for the screening experimental campaigns to be sufficiently parsimonious, however identifying the **main effects** of the factors on the response
  - **full factorial design** to obtain enough information in such a way as to identify both **main effects and factors interactions** on the response
  - **central composite design** to carry out a sufficiently wide experimentation and find the optimal conditions on the system under study (where **main effect, factor interaction and quadratic effects** are identified)

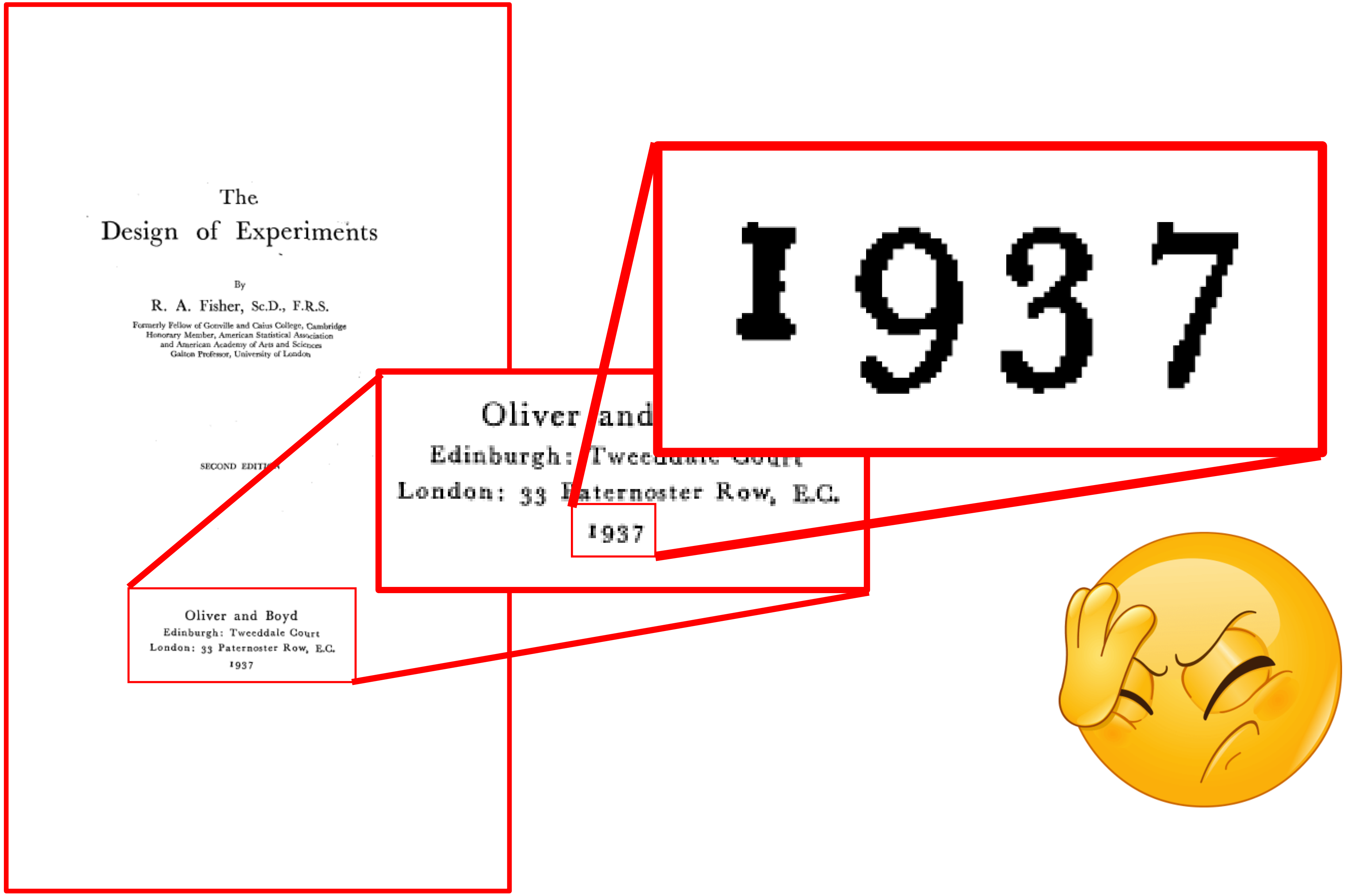
Appendix – Irregular experimental domains

- The experimental domain is usually regular and in form of an (hyper-) cube
- Experiments may not be allowed in some portions of the experimental domain due to unwanted process complications, excessive costs, high energy consumption, different process mechanisms, highly toxic by-products, etc...
  - an irregular domain is present
- Methodologies for handling an **irregular experimental domain**:
  - to **shrink the factor ranges** so that the region becomes regular
    - the hatched area overlooks certain parts of the domain
  - to make use of a **D-optimal design**
    - since an irregular region is more complex than a regular one, the former demands more runs than the latter
- In the problem formulation, an irregular experimental region may be defined by specifying ***linear constraints***, which are linear functions of the quantitative factors that describe the region to be excluded by the experimental domain





DoE, a new discipline…



The
Design of Experiments

By
R. A. Fisher, Sc.D., F.R.S.
Formerly Fellow of Gonville and Caius College, Cambridge
Honorary Member, American Statistical Association
and American Academy of Arts and Sciences
Galton Professor, University of London

SECOND EDITI

Oliver and Boyd
Edinburgh: Tweeddale Court
London: 33 Paternoster Row, E.C.
1937

… per sempre a fianco a me!