Definition

Chemometrics:

The combination of basic science and multivariate analysis (MVA)

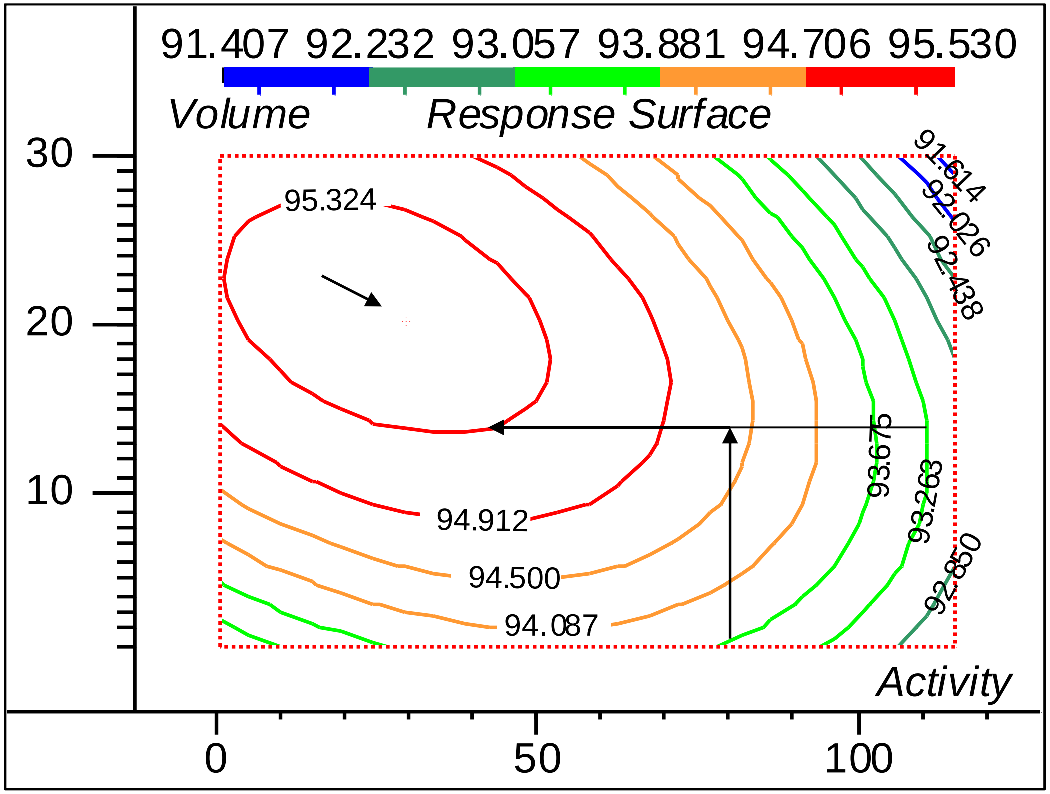
Good chemometrics = good basic science + good MVA

Multivariate analysis (MVA):

Screening and optimization design of experiments (DOE) and multivariate modeling algorithms to extract trends and measure relationships

Knut Dyrstad

Optimumimimum



One factor at a time (OFAT) – Optimum purity cannot be identified

Purpose of multivariate analysis

Identify critical formulation and process variables; these are variables whose levels must be tightly controlled to maintain product stability or performance

Provide limits in which variations in the levels of formulation and process variables have minimal or no effect on product stability or performance (Design Space)

Identify and characterize the interactions among formulation and process variables

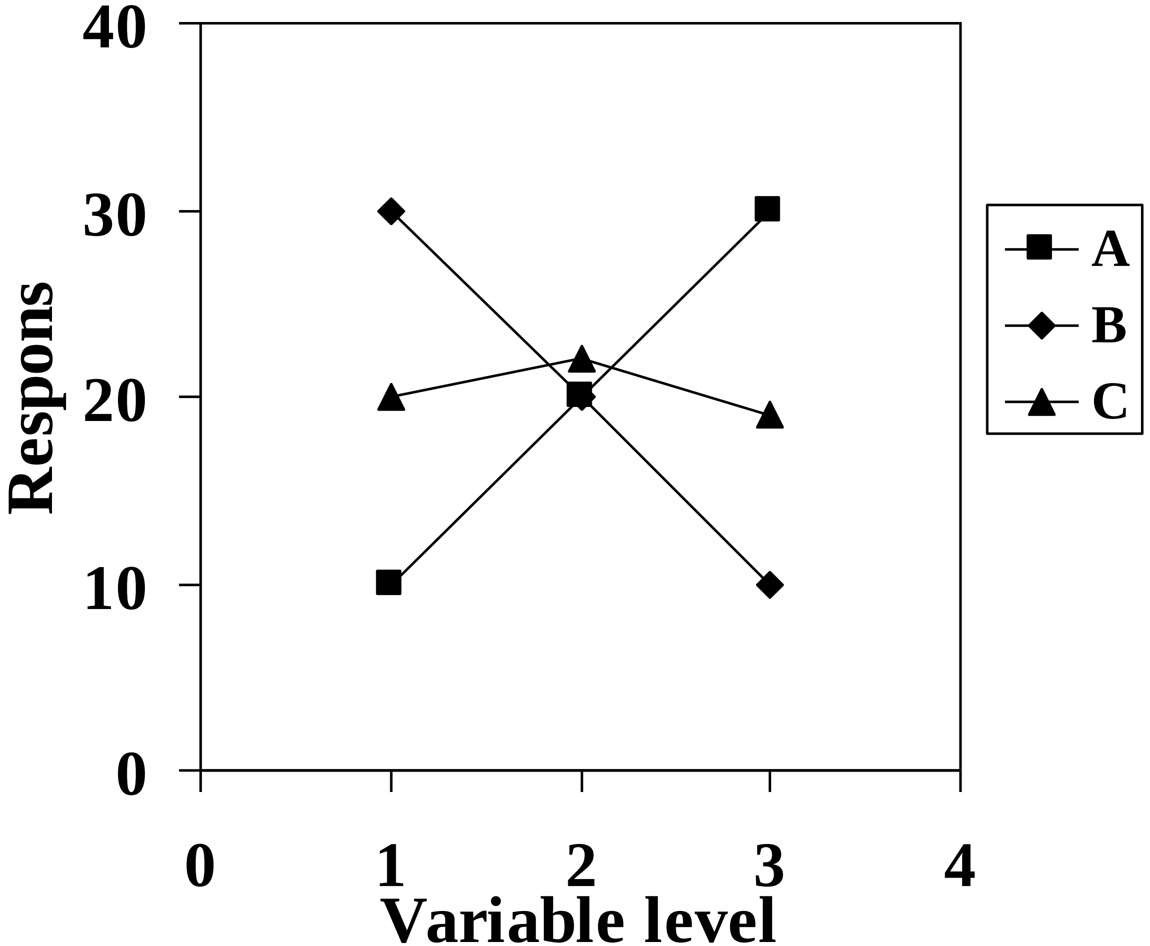
Confirm that the formulation or process is not operating near a design space cliff or discontinuity.

With process means any system where the quality of the output (Ys) is dependent on the adjustment (and control) of input variables (Xs as time, temp, concentrations, etc) for any chem/pharm production process, organic synthesis, HPLC analytical method, etc.

UiO

- Lineære sammenhenger

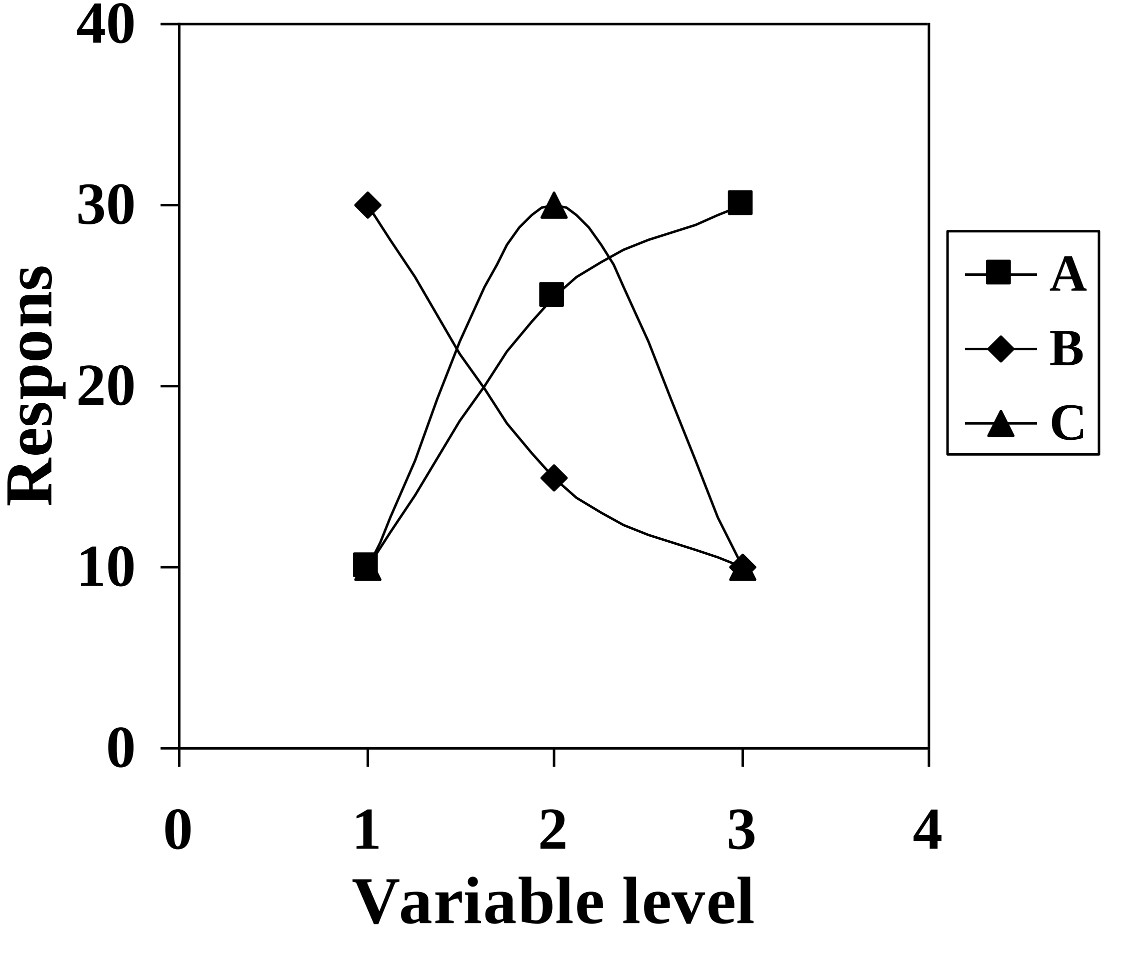
- R = $ß_0 + ß_1 V$
- Y = $ß_0 + ß_1 A$ der $ß_1$=stigningstallet



UiO

- Univariate ikke-lineære relasjoner

- $\log(Y)=ß_0+ß_1A$ eller
- $Y=ß_0+ß_1A+ß_2A^2$



X (conc) mot Y (respons):

- Stigningstallet (regresjonskoeffisient), 3.56

$$b = \frac{\sum_i \{(x_i - \bar{x})(y_i - \bar{y})\}}{\sum_i (x_i - \bar{x})^2}$$

- Interceptet, -1.44

$$a = \bar{y} - b\bar{x}$$

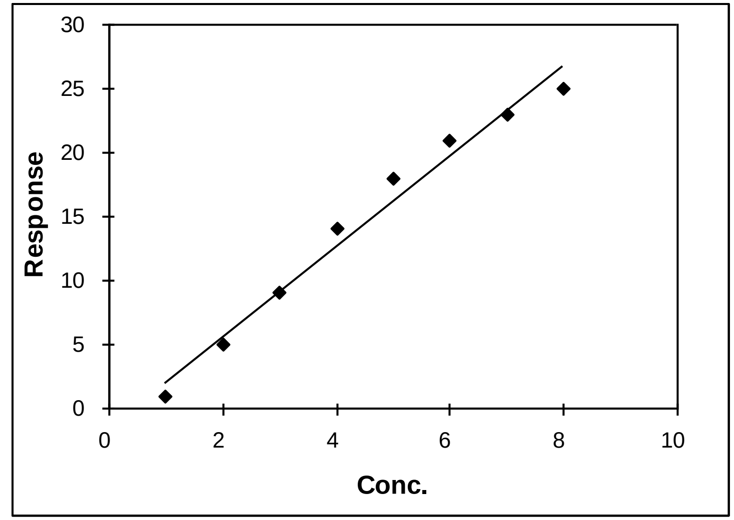
- Standardaviket rundt linja, 1.38

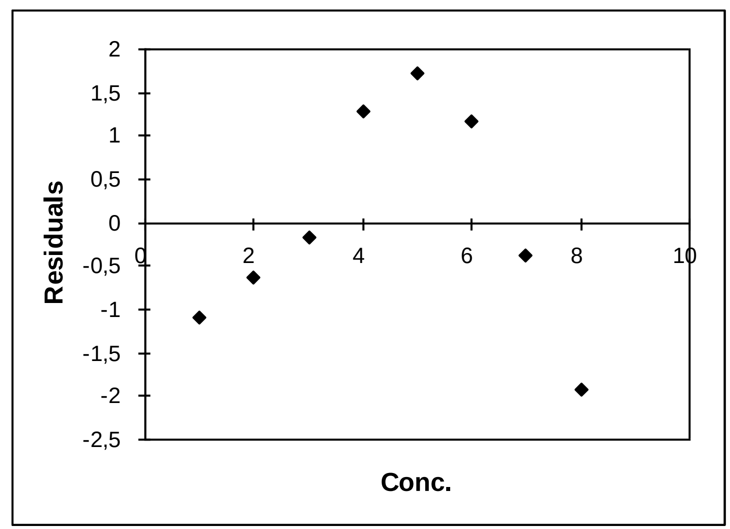
$$s_{y/x} = \left\{ \frac{\sum_i (y_i - \hat{y}_i)^2}{n-2} \right\}^{\frac{1}{2}}$$

- korrelasjonskoeffisienten (-1 til 1), 0.99

$$r = \frac{\sum_i \{(x_i - \bar{x})(y_i - \bar{y})\}}{\{[\sum_i (x_i - \bar{x})^2][\sum_i (y_i - \bar{y})^2]\}^{\frac{1}{2}}}$$

- Residualene (se etter trender)



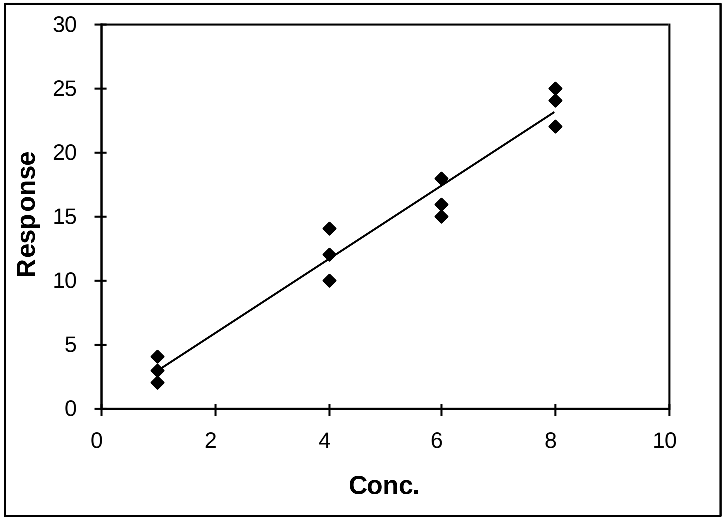


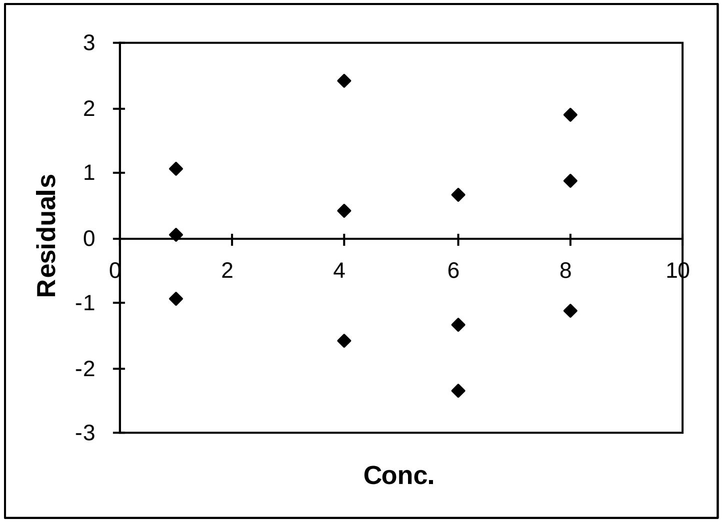
- Univariate relasjoner, X (conc) mot Y (respons):
- Stigningstallet, 2.88
- Interceptet, 0.06
- Sy/x, 1.55
- r, 0.98

- Sb = 0.17, Sa = 0.93

$$s_b = \frac{s_{y/x}}{\{\sum_i (x_i - \bar{x})^2\}^{\frac{1}{2}}}$$

$$s_a = s_{y/x}\left\{\frac{\sum_i x_i^2}{n\sum_i (x_i - \bar{x})^2}\right\}^{\frac{1}{2}}$$





$$\bar{x} = \sum_i x_i/n \,, \qquad s = \sqrt{\sum_i (x_i - \bar{x})^2/(n-1)} \,, \qquad \mu = \bar{x} \pm t(s/\sqrt{n})$$

$$y = bx + a$$

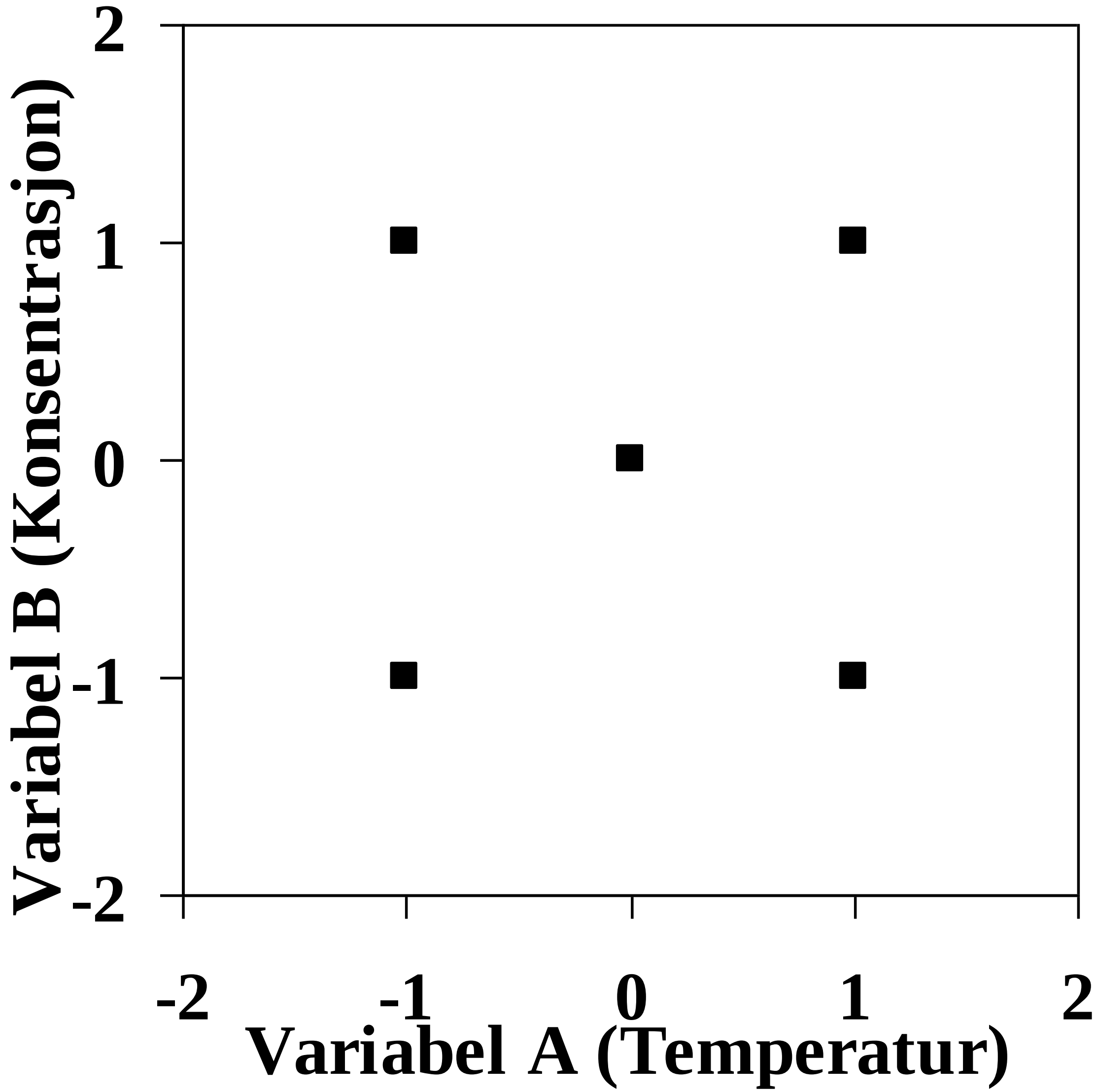
$$b = \frac{\sum_i \{(x_i - \bar{x})(y_i - \bar{y})\}}{\sum_i (x_i - \bar{x})^2} \,, \quad a = \bar{y} - b\bar{x} \,, \quad s_{y/x} = \left\{ \frac{\sum_i (y_i - \hat{y}_i)^2}{n-2} \right\}^{\frac{1}{2}}$$

$$s_b = \frac{s_{y/x}}{\{\sum_i (x_i - \bar{x})^2\}^{\frac{1}{2}}} \,, \quad s_a = s_{y/x} \left\{ \frac{\sum_i x_i^2}{n \sum_i (x_i - \bar{x})^2} \right\}^{\frac{1}{2}} \,, \quad r = \frac{\sum_i \{(x_i - \bar{x})(y_i - \bar{y})\}}{\{[\sum_i (x_i - \bar{x})^2][\sum_i (y_i - \bar{y})^2]\}^{\frac{1}{2}}}$$

$$s_{x_0} = \frac{s_{y/x}}{b} \left\{ 1 + \frac{1}{n} + \frac{(y_0 - \bar{y})^2}{b^2 \sum_i (x_i - \bar{x})^2} \right\}^{\frac{1}{2}} \,, \quad s_{x_0} = \frac{s_{y/x}}{b} \left\{ \frac{1}{m} + \frac{1}{n} + \frac{(y_0 - \bar{y})^2}{b^2 \sum_i (x_i - \bar{x})^2} \right\}^{\frac{1}{2}}$$

- 2-nivås faktorielt design
- Hvert punkt = et eksperiment
- Variable kodes alltid som -1 (lav) og 1 (høy)
- A og B lineær
- Senterpunkter:
  - sjekke linæritet
  - variasjon

- Y=b0+b1A+b2B+b3AB



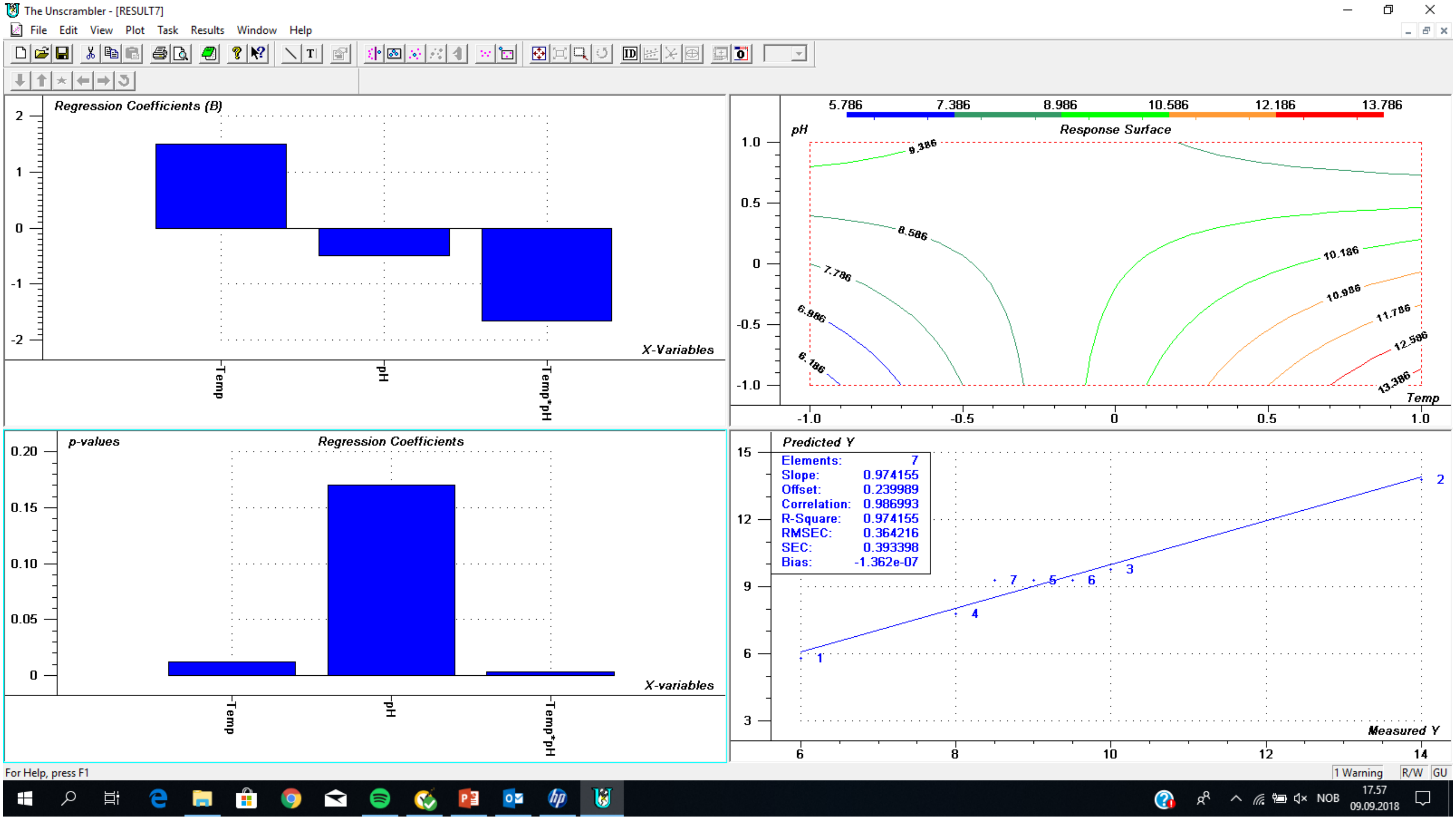
- Screening - 2 nivåes design
  - **Fullt faktorielt design**, formel: 2^n
  - Effekter beregnes vha Yates algoritme for ortogonale design (= 2^n)
  - Eksempel: Effekten av temperatur og pH på utbytte av en kjemisk reaksjon.

Temp.	pH	Temp x pH	Utbytte
-1	-1	1	6
1	-1	-1	14
-1	1	-1	10
1	1	1	8

- Normalisert effekt av temp: (-6+14-10+8)/4 = 1.5
- Normalisert effekt av pH: (-6-14+10+8)/4 = -0.5
- Normalisert effekt av samspill: (+6-14-10+8)/4 = -2.5

– konklusjon: Øke temperatur og senke pH kraftig øker utbytte (mg/ml).

- Eksempel fra Unscrambler

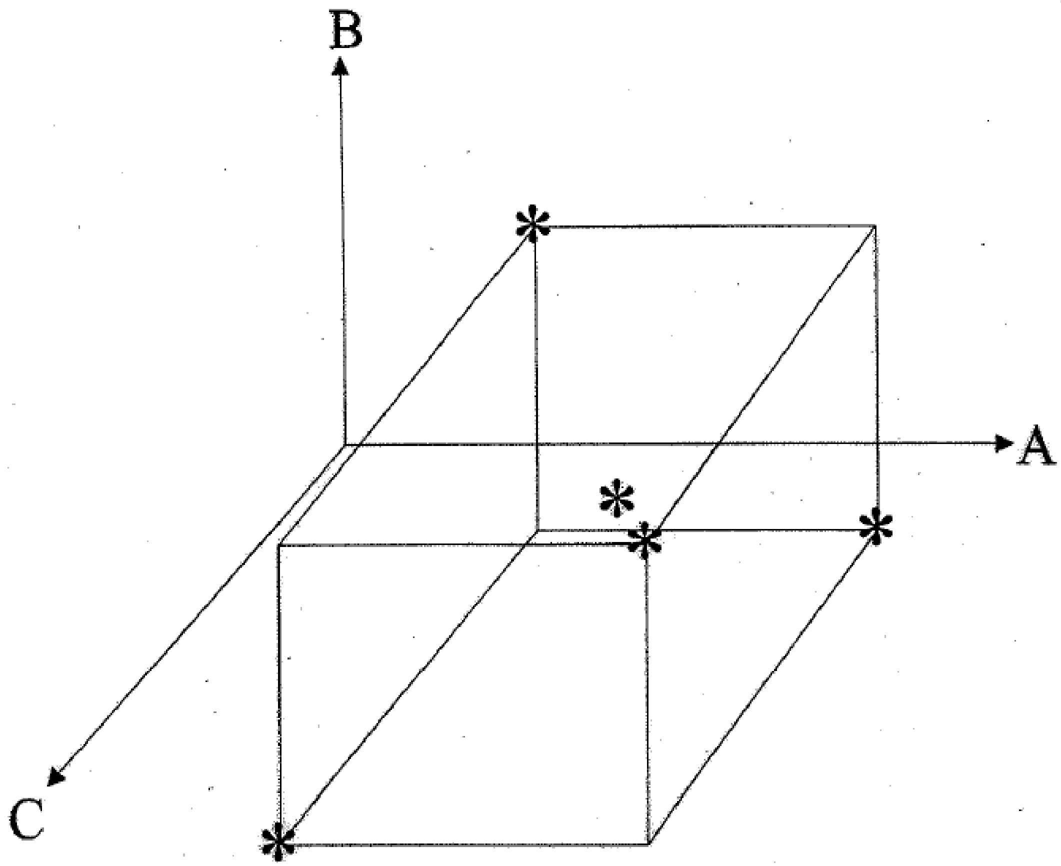


Formler for normalisering av variabel

1. Normalisert verdi (-1, 1): ((Xn - mean X)/(max X - min X)) x 2

2. Normalsert verdi ved bruk av standard avviket: Xn / STD(Xs)

UiO



Hjørnestjernene angir det reduserte designet

UiO

- Screening - 2 nivåes design
  - **Fraksjonelt faktorielt design**, formel: 2^{n-k}, gjelder for n>2.
  - Effekter beregnes vha Yates algoritme for ortogonale design (= 2^{n-k})
  - Eksempel: Effekten av temperatur, pH og mengde antioksidant på stabiliteten av virkestoffet etter lagring i 1 måned.

Temp	pH	Antioksidant	T x pH	T x A	pH x A	% recovery
-1	-1	1	1	-1	-1	98
1	-1	-1	-1	-1	1	80
-1	1	-1	-1	1	-1	88
1	1	1	1	1	1	78

- Normalisert effekter: Temp = -7, pH = -3 og antioksidant = 8.

  - konklusjon 1: Virkestoffet er følsomt for temperatur, pH bør reduseres og høy mengde antioksidant reduserer degradering.
  - konklusjon 2: Hoved- og samspillseffekter er blandet sammen.

- Oppløsning VI
  - Alle hoved- og samspillseffekter er ublandet.
- Oppløsning V
  - Alle hovedeffekter og 2-faktorsamspill er ublandet.
- Oppløsning IV
  - Hovedeffekter er blandet med 3-faktorsamspill og 2-faktorsamspill er blandet med andre 2-faktorsamspill.
- Oppløsning III
  - Hovedeffekter er blandet sammen med 2 faktorsamspill.

UiO

- Optimalisering (design og effekter utledes vha PC!)
  - Fullt faktorielt design, 3^n
  - Central composite
  - Box-Behnken

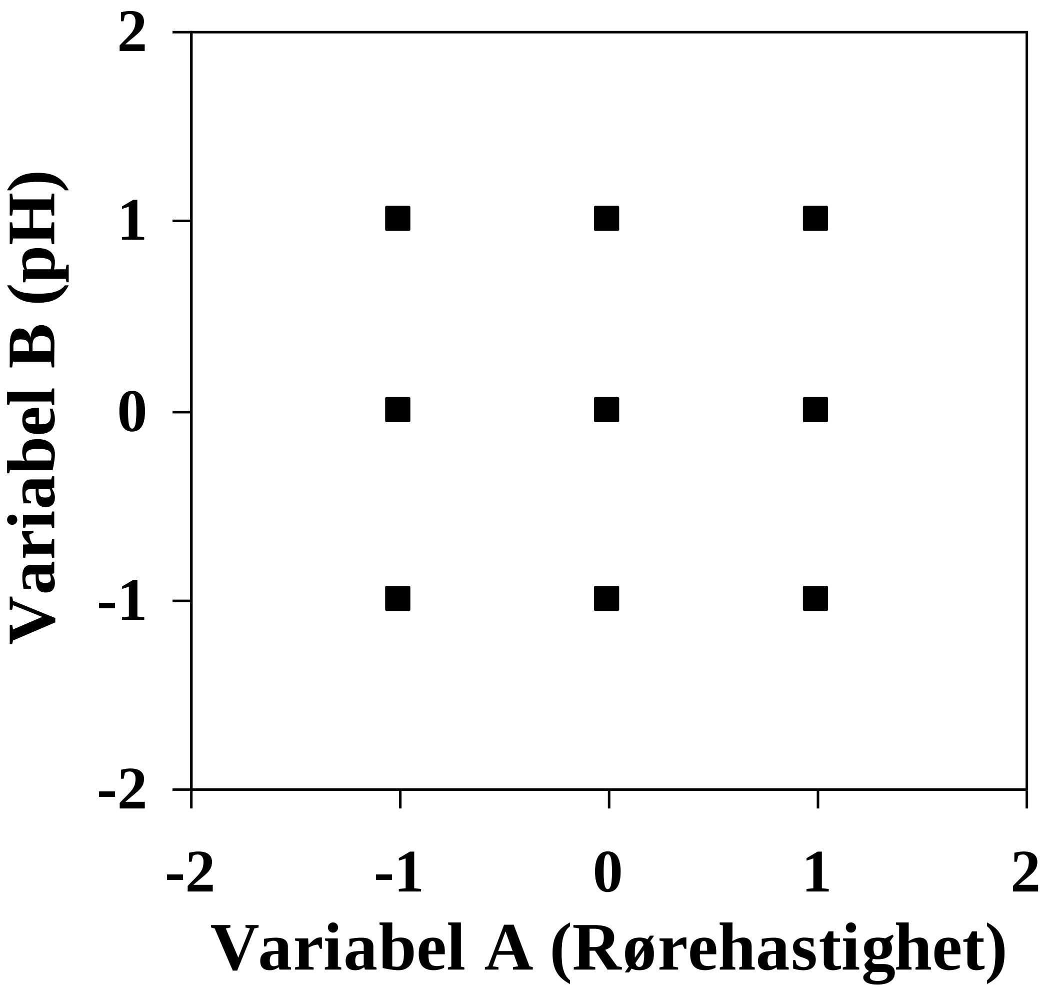
- Mixed Level
  - K^m x L^n, design etter behov

- Mixture design (design og effekter utledes vha PC!)
  - Simplex lattice

UiO

- Faktorielt Optimaliserings-design
- A og B ulineær
- 3 nivåer

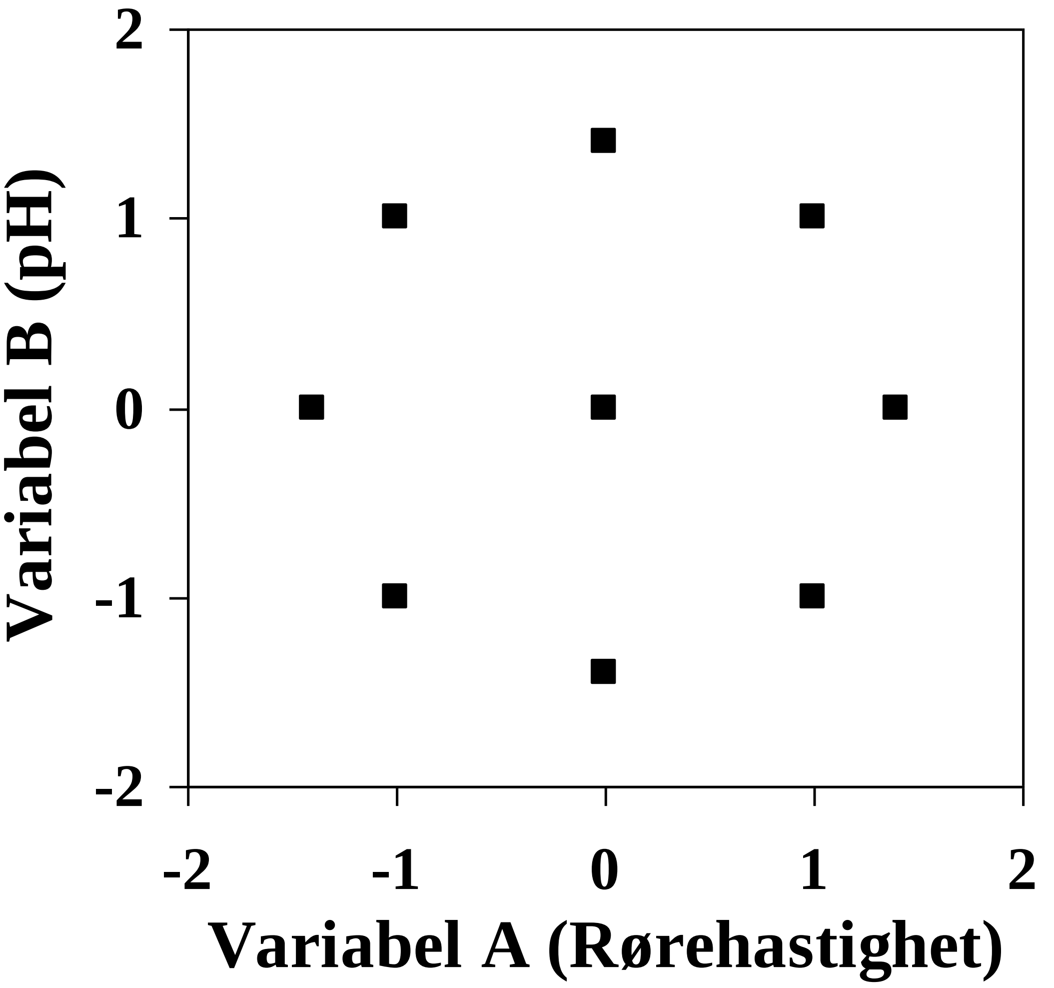
- Y=b0+b1A+b2B+ b3AA+b4BB+b5AB +b6AAB+b7ABB



UiO

- Central composite design (CCD)
- Optimalisering
- A og B ulineær
- 5 nivåer

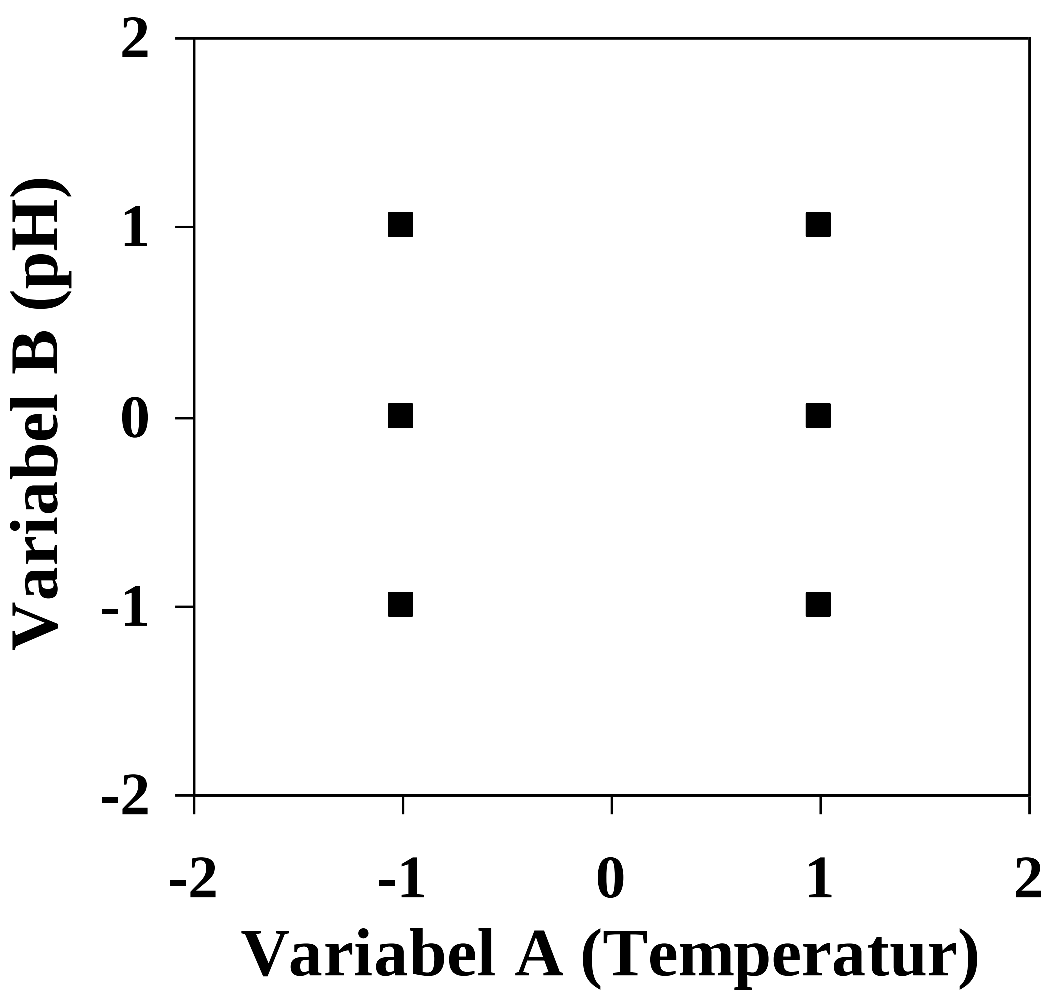
- Y=b0+b1A+b2B+ b3AB+b4AA+b5BB+ b6AAA+b7BBB



UiO

- Faktorielt design
- Mixed level
- A lineær
- B ulineær

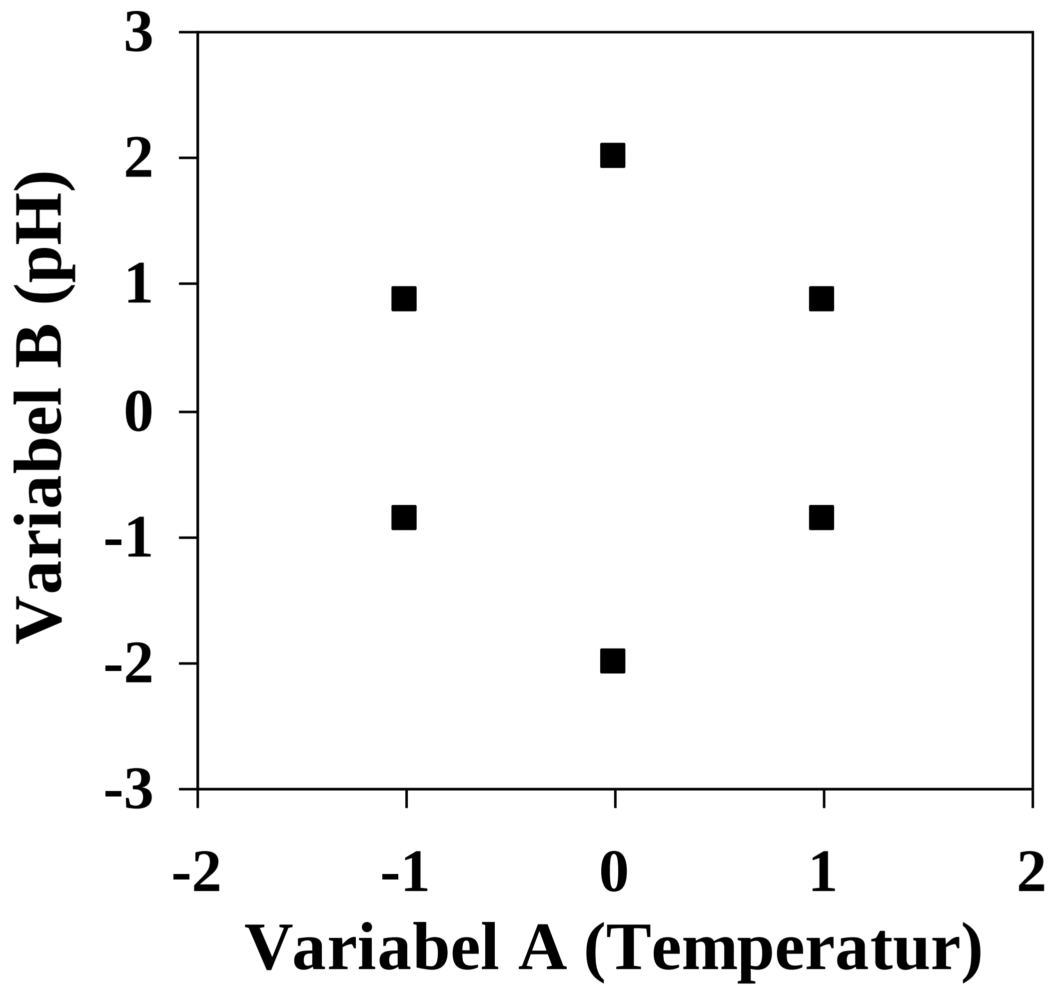
- Y=b0+b1A+b2B+ b3AB+b4BB



UiO

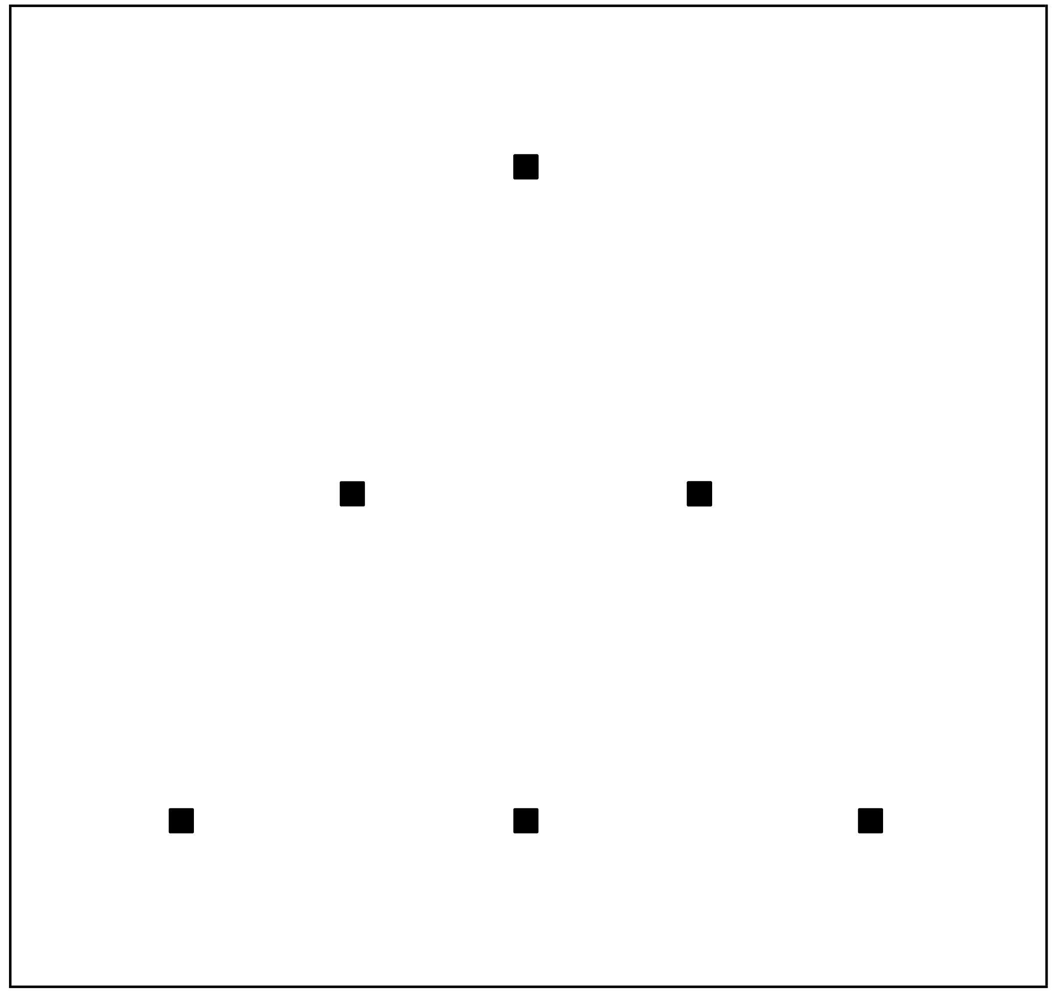
- Faktorielt design
- Mixed level
- A lineær
- B ulineær

- Y=b0+b1A+b2B+b3AB+b4BB+b5BBB

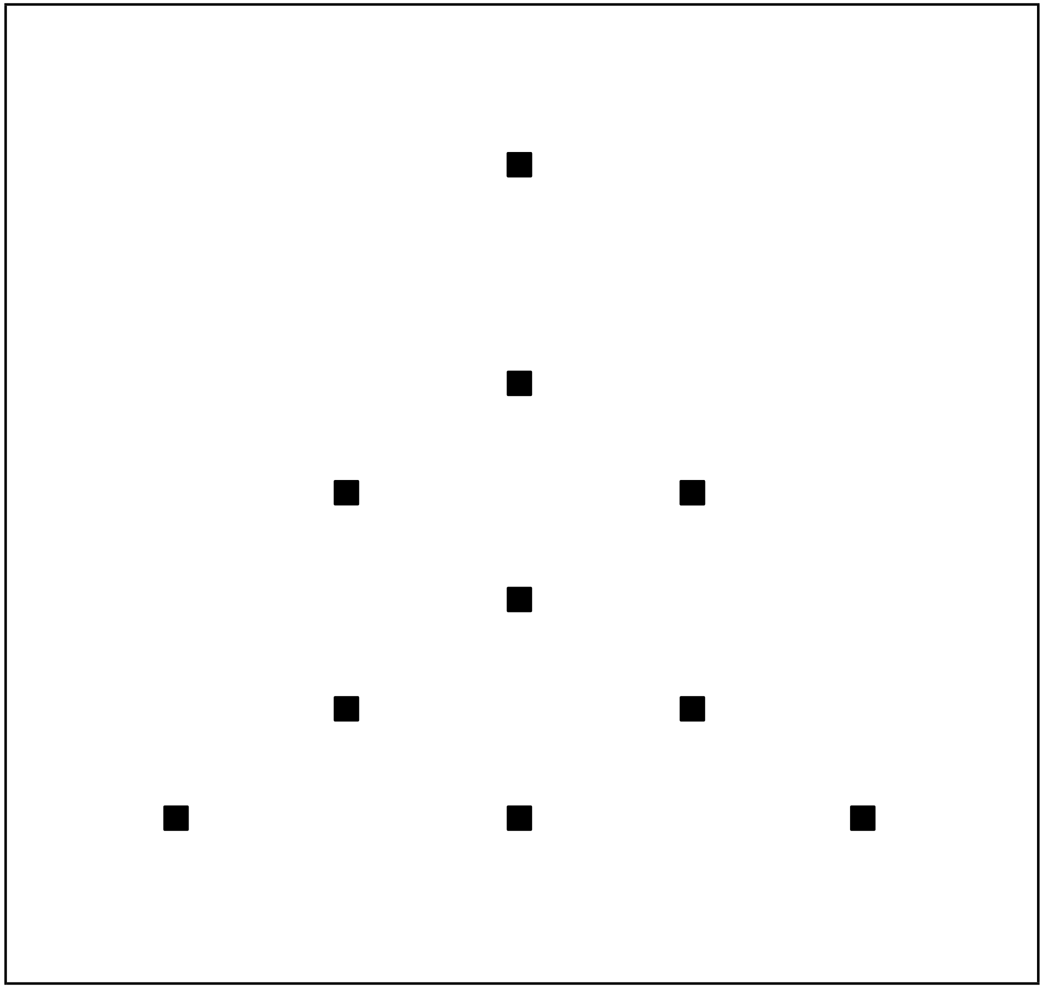


UiO

- Mixture design
- Simplex-Lattice
- 2. ordens design
- Punktense verdier for X-variablene leses av på vektoren fra 0-linjen til variabel til motliggende toppunkt



- Mixture design
- Augmented Simplex-Lattice
- 2. ordens design
- kombinasjoner av alle 3 variable
- Punktenes verdier for X-variablene leses av på vektoren fra 0-linjen til toppunktet for respektive variable



The experimental process

1) Set the objectives

2) Brainstorming – possible causes of the problem and factors that influence the product or process
   - Prioritize the factors.
   - Determine the actual High/Low values

3) Evaluate whether uncertainty in the production/analysis might affect the interpretation.
   - Possible introductory experiments

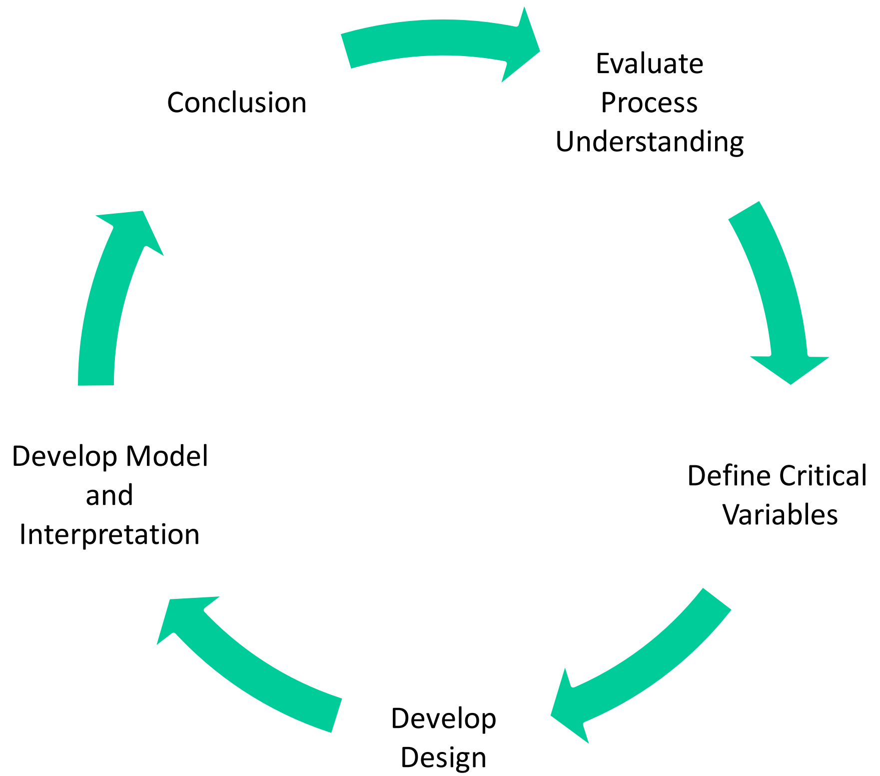
4) Consider probable linear/quadratic/interaction effects and choose a suitable design

5) Running of the experiments and exploration of the design space
   - Randomized sequence
   - Keep all non-design factors constant
   - Register possible differences (equipment, person, time,…)

6) Analysing of data and interpretation of the results
   – creation of the meta-model (~the first model)

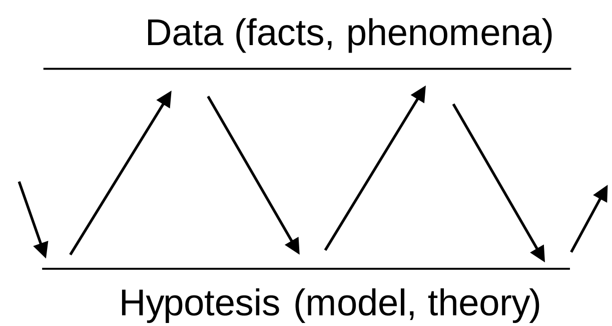
7) Validation of the meta-model and use of it in subsequent steps to optimise the design

The chemometrics process is continuous



Start with screening variables in highly reduced (fractionized) designs and end testing the most critical variables in a targeted optimization design

The strategy leads to improved process understanding (process design space), process reliability and robustness, and is also the needed first step to start a QbD process to simplify QC



DoE strategy - Exploring the design space

Step 1

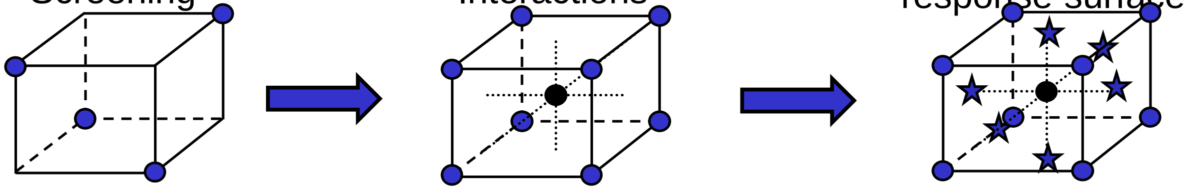
Screening

Step 2

Interactions

Step 3

Central composite for response surface



Output from multivariate modeling

1) Determine important variables and interaction effects

2) Find best/optimal variable settings (response surface)

2) Establishing design space (operating window, Spec, QbD)

3) A measure of robustness and precision (future reproducibility)

4) Deviation from prediction reveals new variables/mechanisms

5) Graphical interpretations of complex relations

6) Summarise all info in data (modelling is the only way)

7) Accelerate the theoretical understanding, -balancing empirical and (theoretical) process understanding is a goal