# An introduction to Design of Experiments

Harald Martens (1) and Frank Westad (2)
(1) Prof.emerit. ITK & founder, Idletechs AS
(2) Prof. II ITK & CSO CAMO Software
harald.martens@ntnu.no, frank.westad@ntnu.no

# WHY???

***Why have only humans white eye balls?***

# WHY???

***Why have only humans white eye balls?***

***Why is the thunder cloud black?***

# WHY???

*Why have only humans white eye balls?*

*Why is the thunder cloud black?*

*Which food dye gives white color?*

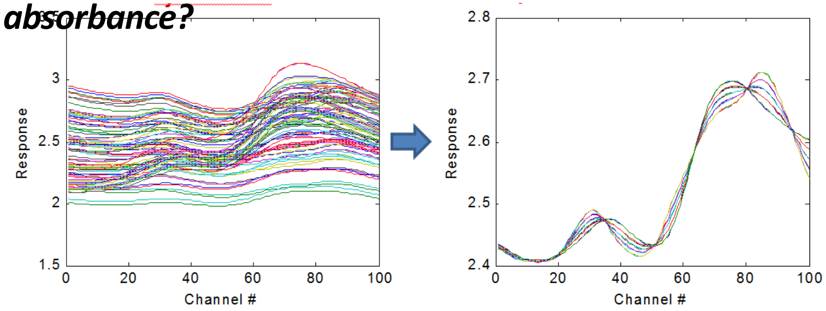
# WHY???

### *Why have only humans white eye balls?*

### *Why is the thunder cloud black?*

### *Which food dye gives white color?*

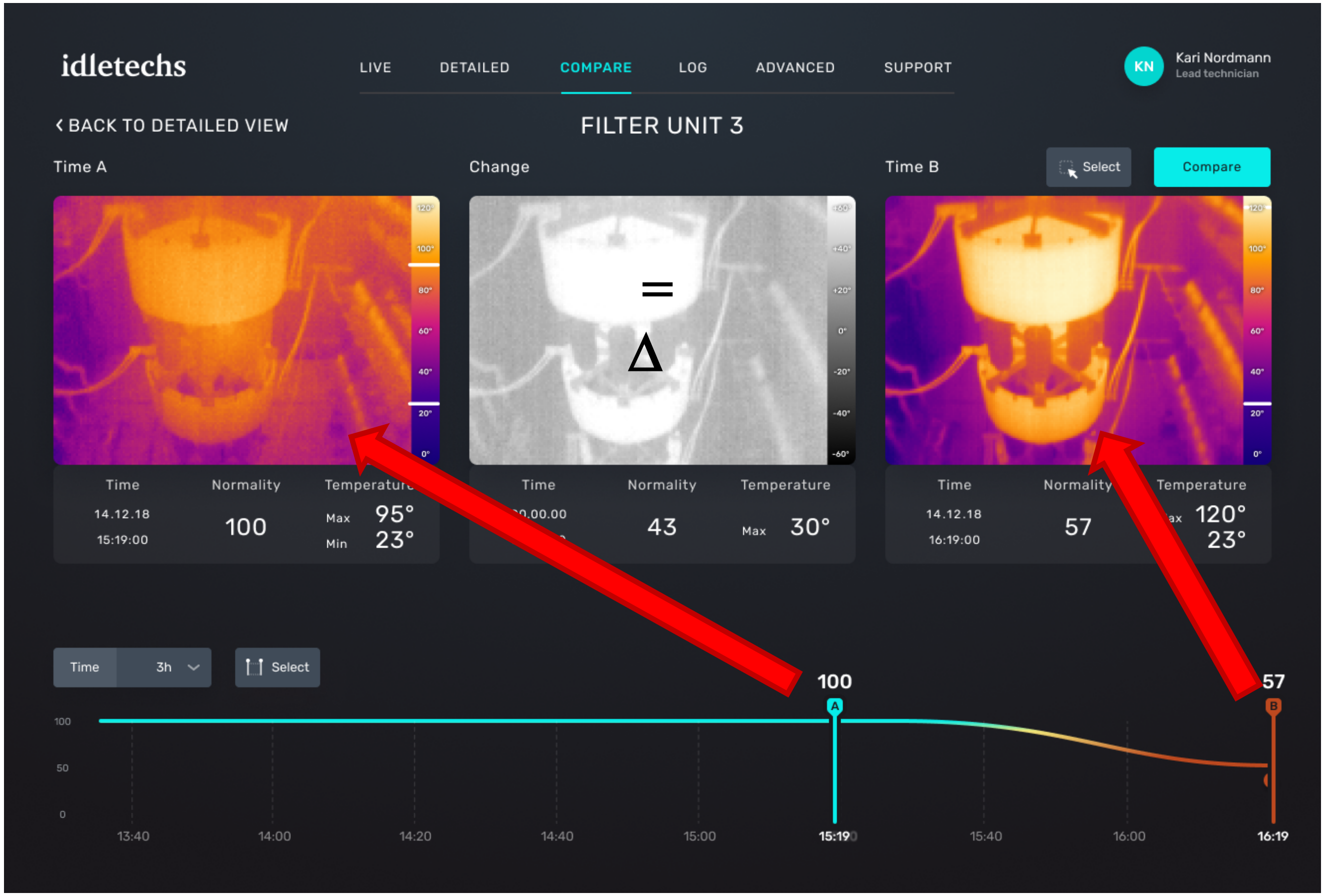
### *How do you separate light scattering from light absorbance?*



# Contents

- Introduction
- Factorial designs
- ANalysis Of VAriance (ANOVA)
- Short overview of design types

# Making quantitative data understandable for ordinary people: e.g. XAI for thermal analysis of high-power electrical equipment

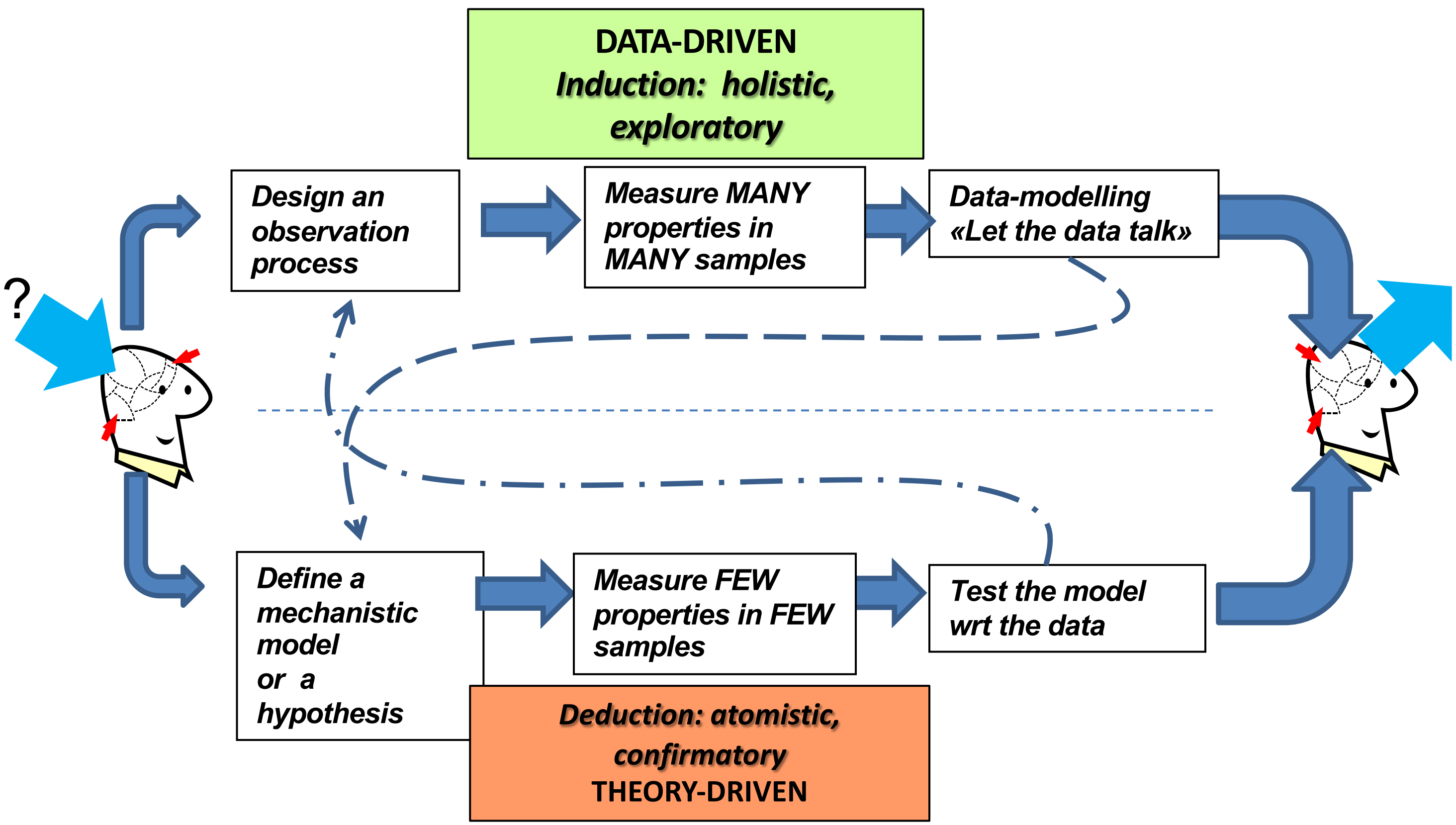



idletechs

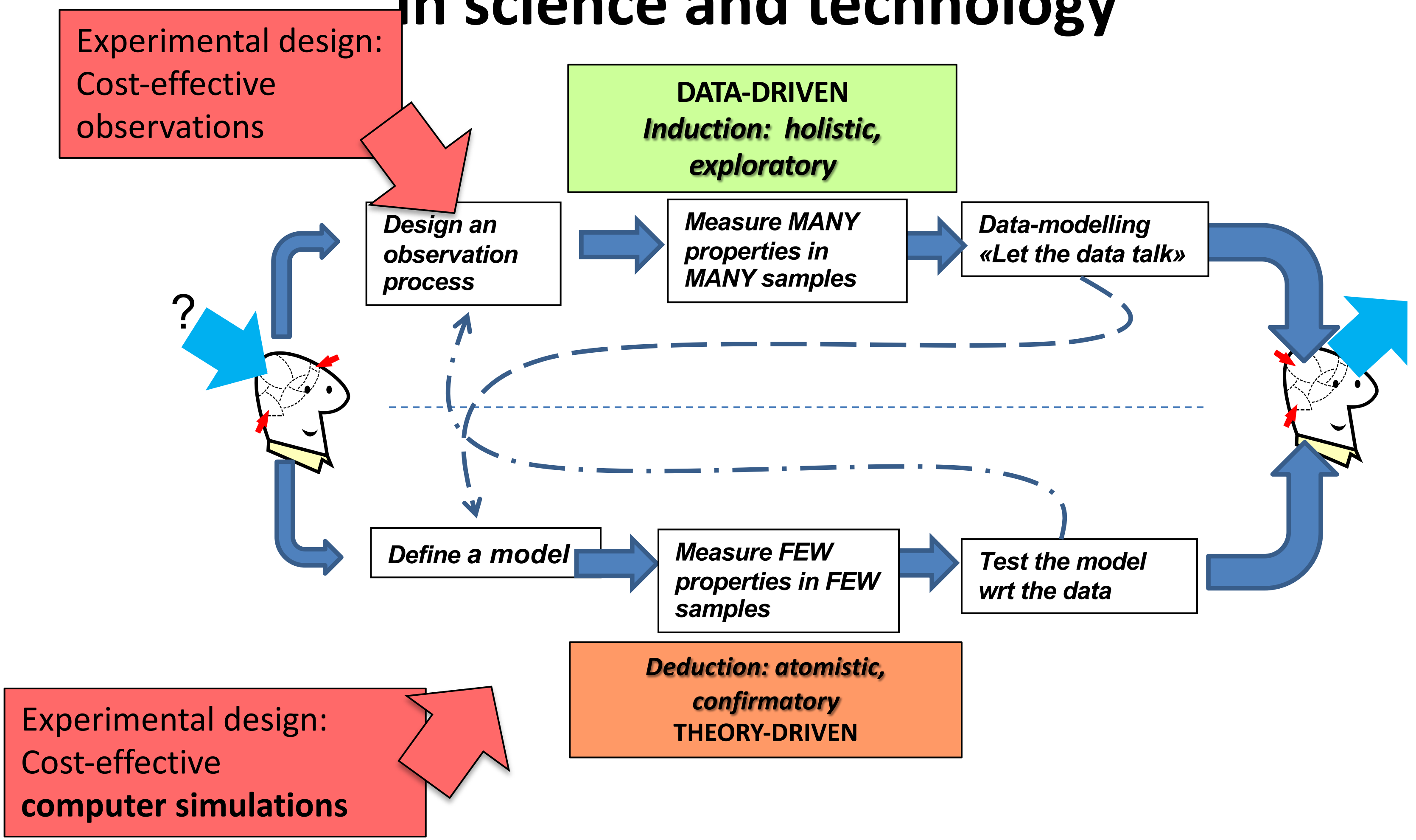
# Why investigate?

- Why controlled experiments?
  - To understand causal effects, in order to control or optimize an effect.
  - To know/understand:
    - 1) be convinced,
    - 2) have good reason to be convinced,
    - 3) conviction actually true
  - Mathematical models: Efficient summaries of data
  - Statistical assessments: guard against our confirmation bias / wishful thinking
- Learning from data: Hypothesis-driven vs Data driven modelling: Deduction vs Induction

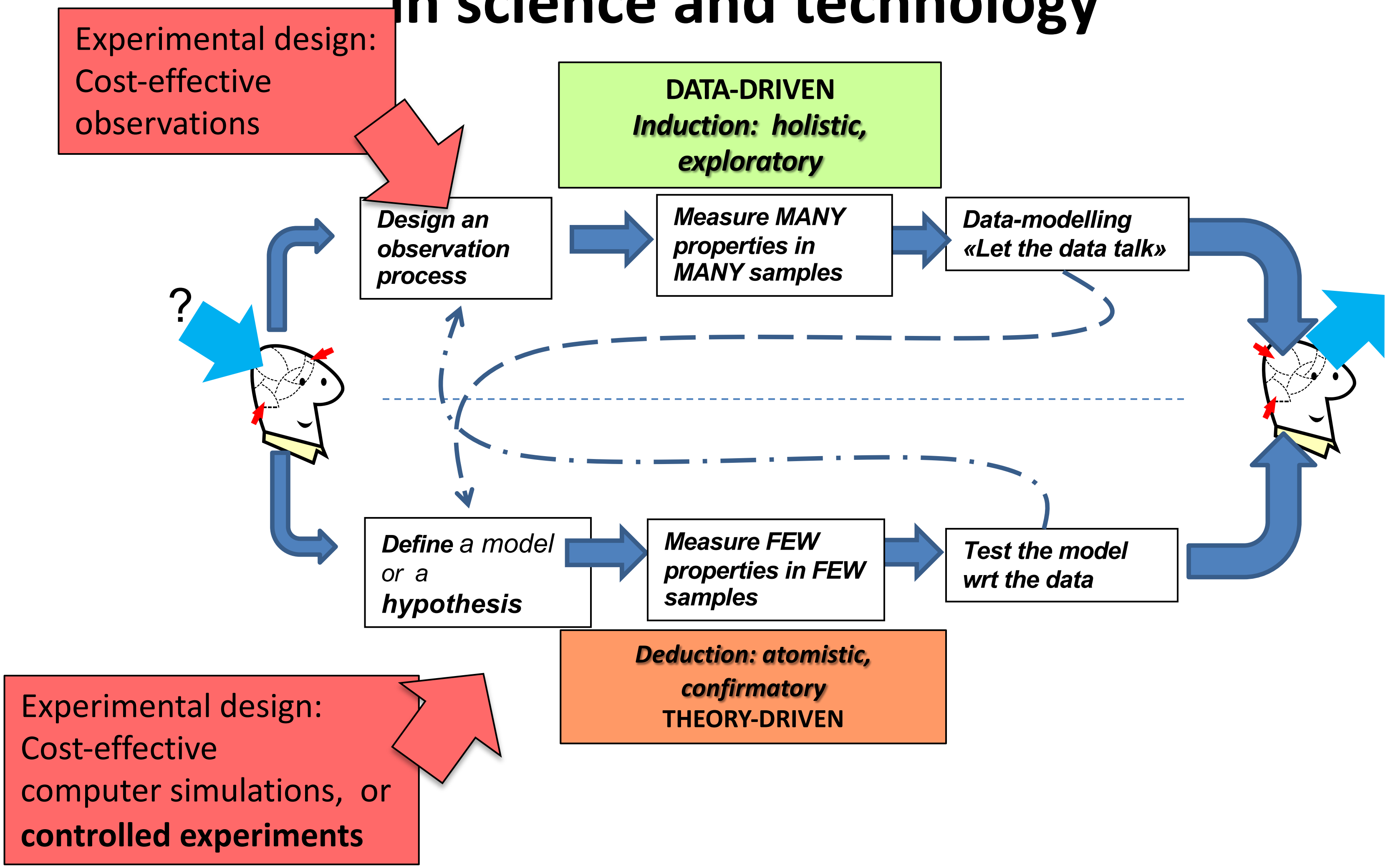
# Two roads from question to answer in science and technology



# Two roads from question to answer in science and technology



# Two roads from question to answer in science and technology



# Statistics: Generally about correlations, not causality

- Cybernetics: Use small data + Kalman-F. continuously to compensate
- Big Data Cybernetics:  Use big data to compensate & get understanding

# Statistics:
# Generally about correlations, not causality

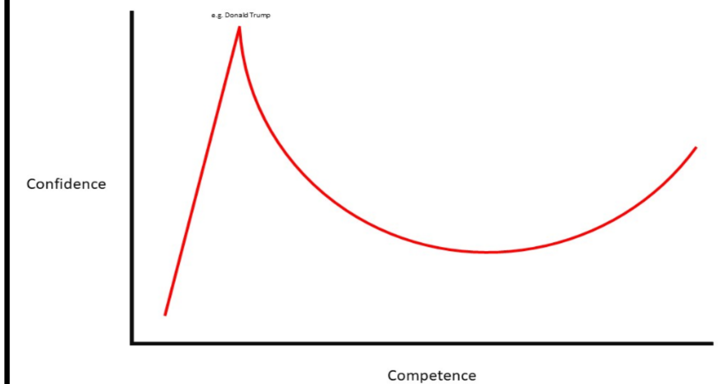
- Cybernetics: Use small data + Kalman-F. continuously to compensate
- Big Data Cybernetics:  Use big data to compensate & get understanding

- Causal modelling and statistical fitting to data!
- Forgotten error sources:
  - Confirmation bias: Wishful thinking.
  - Alias errors due to incomplete modelling:  ‘lurking variables’ = unmodelled effects

# People often do the right things for the wrong reasons

**Tacit knowledge: Important!!**
**But:**
Inflexible,
Difficult to communicate,
Sometimes wrong

The classical definition of **knowledge**:
believed, justified and true.
https://en.wikipedia.org/wiki/Knowledge#Theories_of_knowledge



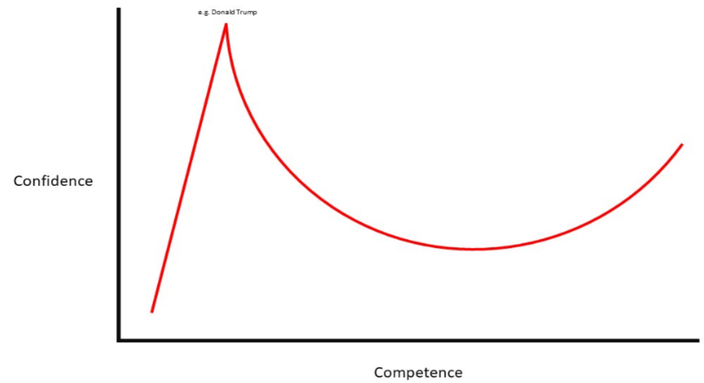

Dunning-Kruger effect
https://en.wikipedia.org/wiki/Dunning%E2%80%93Kruger_effect

# People often do the right things for the wrong reasons

**Tacit knowledge: Important!!**
**But:**
Inflexible,
Difficult to communicate,
Sometimes wrong

The classical definition of **knowledge**:
believed, justified and true.
https://en.wikipedia.org/wiki/Knowledge#Theories_of_knowledge

In Science, Industry & Society:
**Tacit knowledge → Explicit knowledge:**
**To *understand* pays off!**



Dunning-Kruger effect
https://en.wikipedia.org/wiki/Dunning%E2%80%93Kruger_effect

***Understanding*, Comprehension: Ability to think about and to deal adequately with an idea, support intelligent behaviour**
https://en.wikipedia.org/wiki/Comprehension

**Comprehension is an ability to compress data.** https://en.wikipedia.org/wiki/Gregory_Chaitin
***Math models: Efficient compression of data and of knowledge***

**Theory-based, models of mechanisms ---- Observation-based models of correlation patterns**

The informative converse paradox: Windows into the unknown

H. Martens / Chemometrics and Intelligent Laboratory Systems 107 (2011) 124–138

Where we think we know the most,
we make the most mistakes

# Separating different additive causes by incomplete modelling ?

The informative converse paradox: Windows into the unknown

H. Martens / Chemometrics and Intelligent Laboratory Systems 107 (2011) 124–138

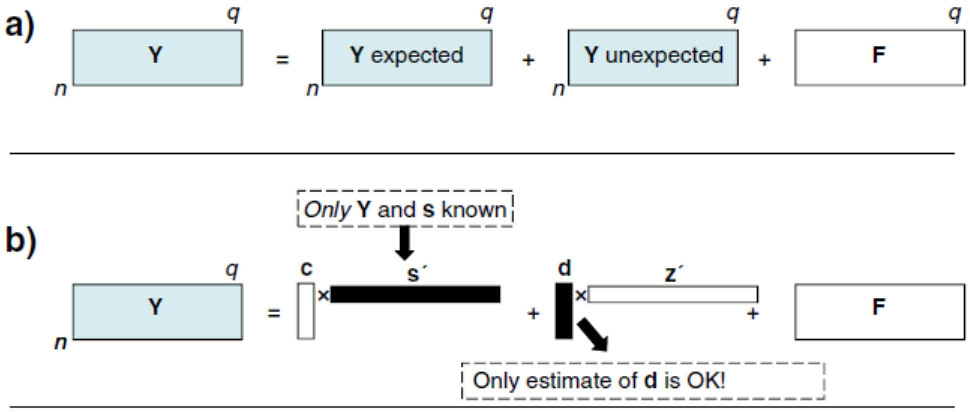
a)

$$\underset{n \times q}{\mathbf{Y}} = \underset{n \times q}{\mathbf{Y}\ \text{expected}} + \underset{n \times q}{\mathbf{Y}\ \text{unexpected}} + \underset{q}{\mathbf{F}}$$

The informative converse paradox: Windows into the unknown

H. Martens / Chemometrics and Intelligent Laboratory Systems 107 (2011) 124–138

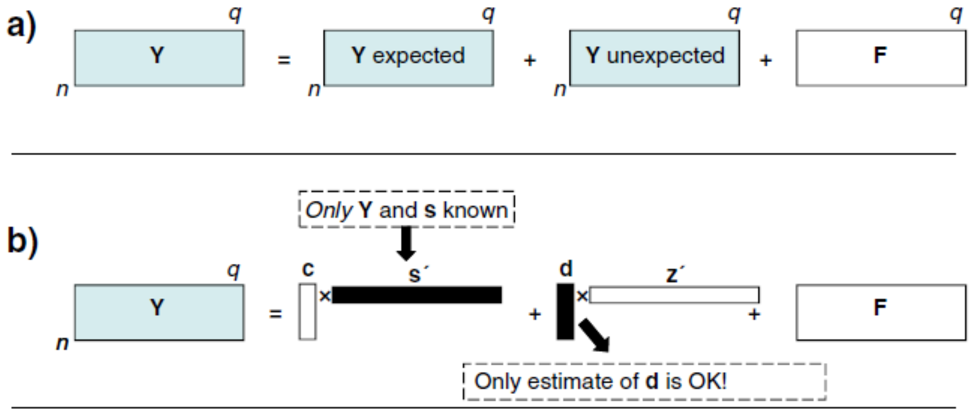
# Separating different additive causes by incomplete modelling ?



The informative converse paradox: Windows into the unknown

H. Martens / Chemometrics and Intelligent Laboratory Systems 107 (2011) 124–138

# Separating different additive causes by incomplete modelling ?




Model: $Y = c\mathbf{s}' + E$

$c = Y\mathbf{s}(\mathbf{s}'\mathbf{s})^{-1}$

$E = Y - c\mathbf{s}'$

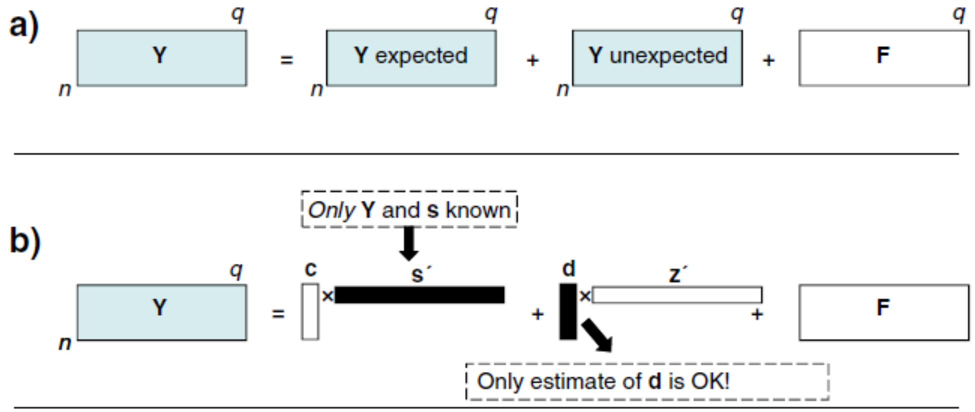
$[d,z] \leftarrow \mathrm{svd}(E)$

$F = Y - c\mathbf{s}' - dz'$

The informative converse paradox: Windows into the unknown

H. Martens / Chemometrics and Intelligent Laboratory Systems 107 (2011) 124–138

# Separating different additive causes by incomplete modelling ?




Model: $Y = c\mathbf{s}' + E$

$c = Y\mathbf{s}(\mathbf{s}'\mathbf{s})^{-1}$

$E = Y - c\mathbf{s}'$

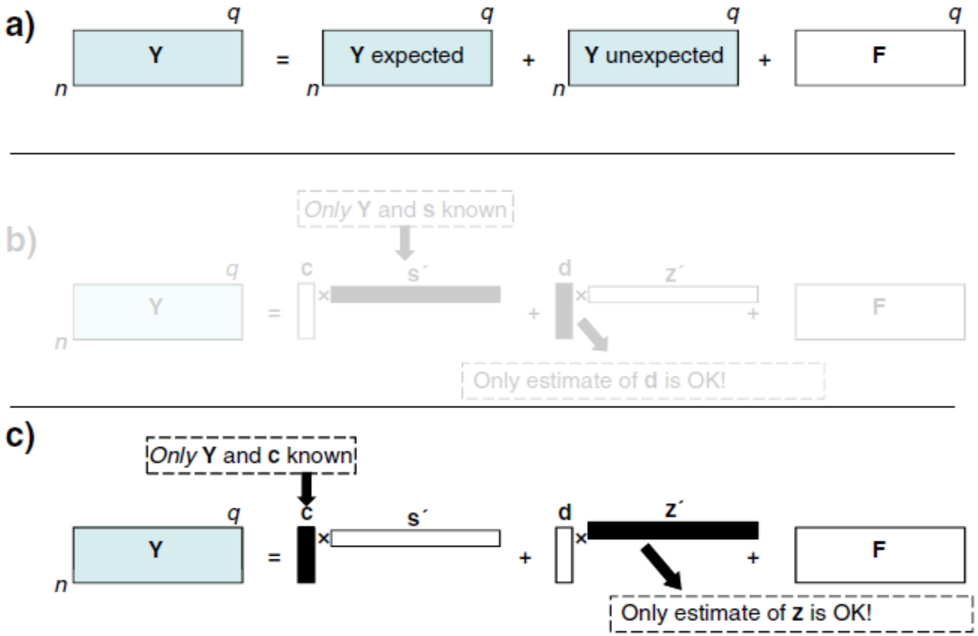
$[d,z] \leftarrow \text{svd}(E)$

$F = Y - c\mathbf{s}' - dz'$

The informative converse paradox: Windows into the unknown

H. Martens / Chemometrics and Intelligent Laboratory Systems 107 (2011) 124–138

# Separating different additive causes by incomplete modelling ?



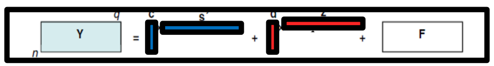

Model: Y=**c**s’ + E

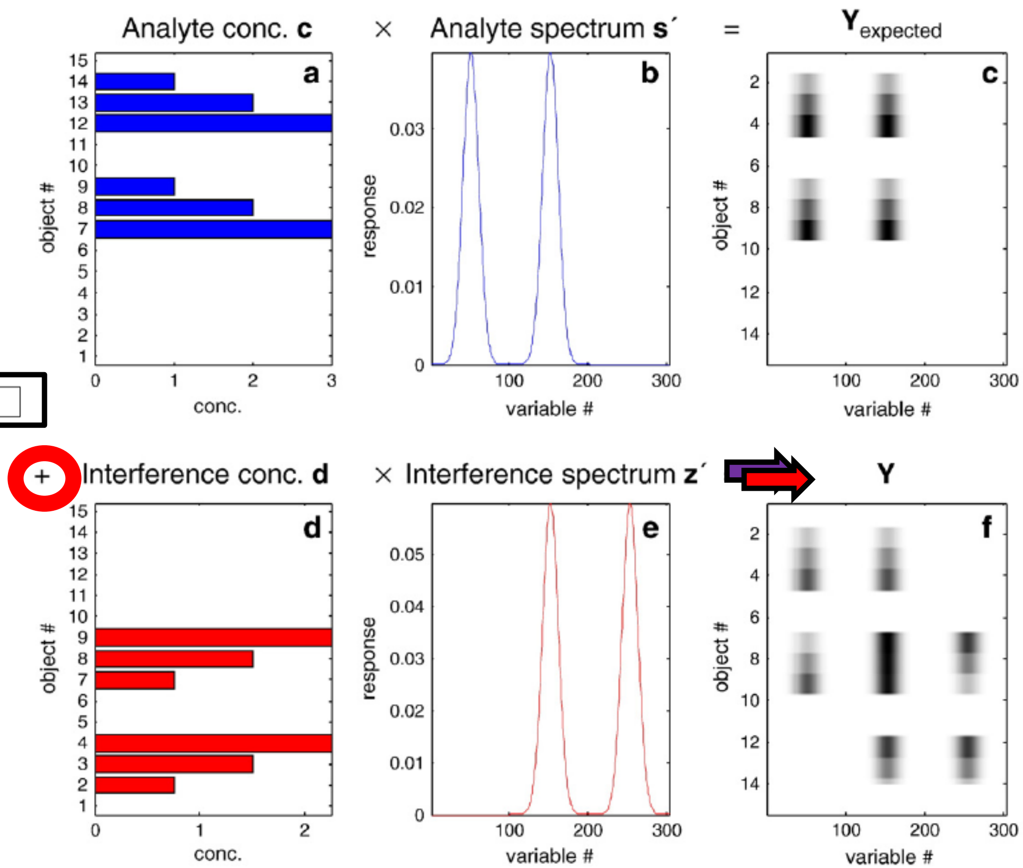
s=(**c**’**c**)$^{-1}$**c**’Y

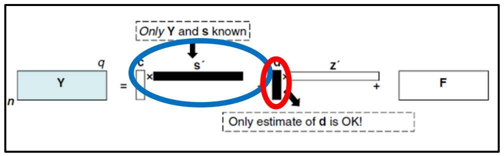
E=Y- **c**s’

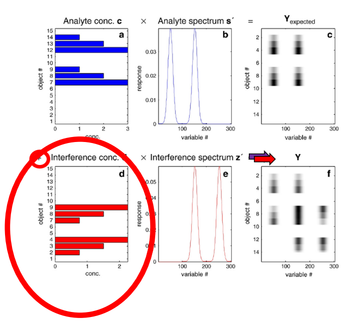
[d,z]← svd(E)

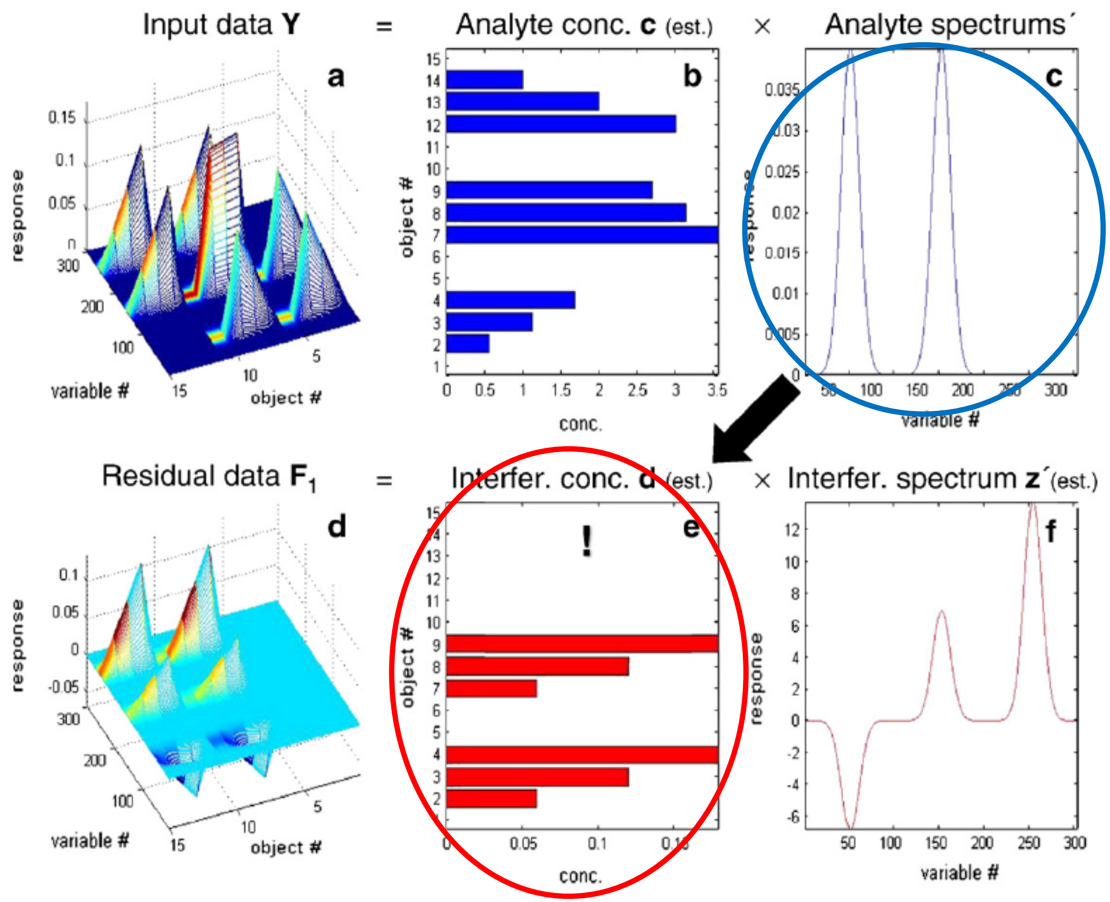
F=Y-**c**s’ – dz’

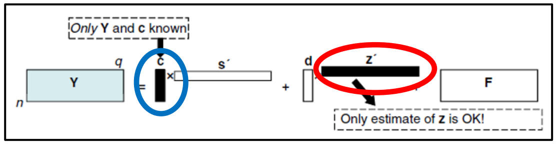

Y
=
c
s´
+
d
z´
+
F
q
n

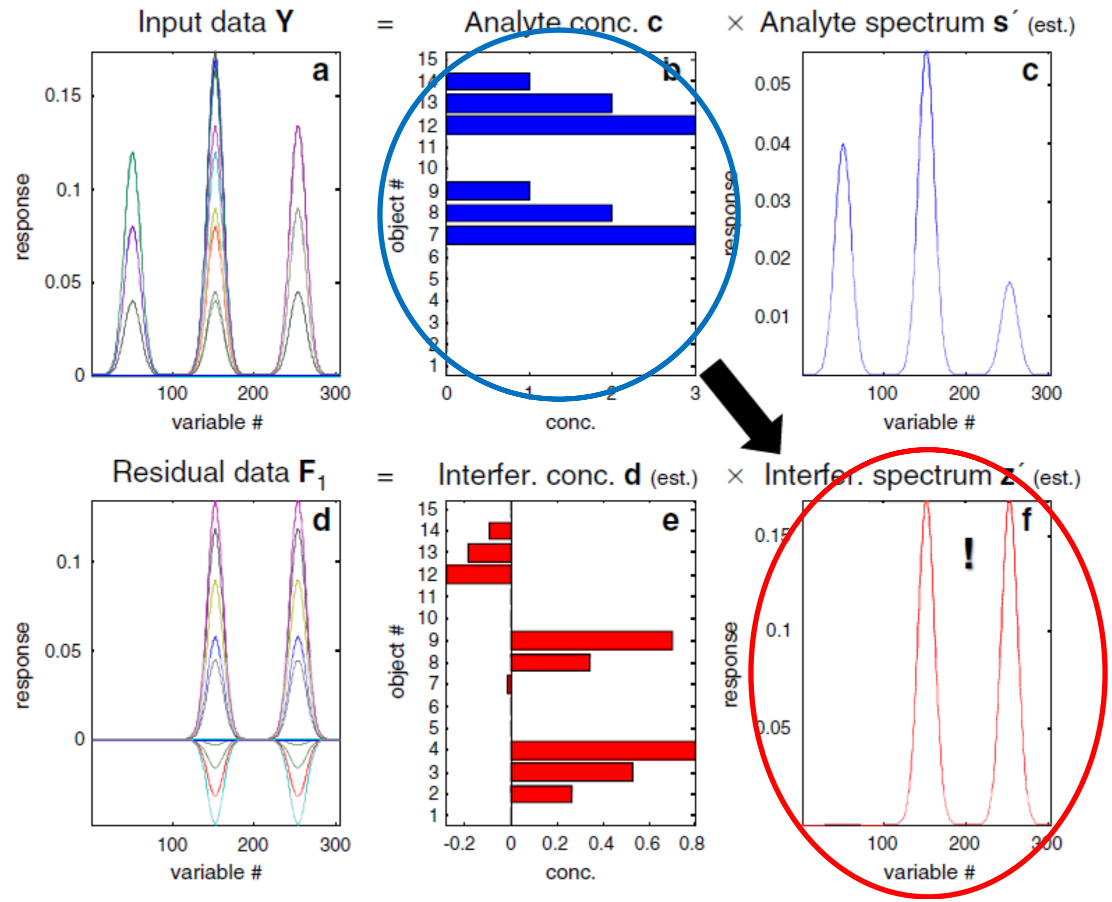

Analyte conc. c
×
Analyte spectrum s´
=
Y expected
a
b
c
object #
conc.
response
variable #
+
Interference conc. d
×
Interference spectrum z´
Y
d
e
f
object #
conc.
response
variable #

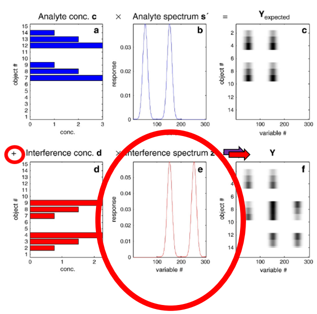

Only Y and s known
Y
q
n
=
c
×
s´
d
z´
+
F
Only estimate of d is OK!


Analyte conc. c
×
Analyte spectrum s´
=
Y expected
a
b
c
object #
response
conc.
variable #
Interference conc. d
×
Interference spectrum z´
Y
d
e
f


Input data Y
=
Analyte conc. c (est.)
×
Analyte spectrums´
a
b
c
response
variable #
object #
conc.
Residual data F1
=
Interfer. conc. d (est.)
×
Interfer. spectrum z´(est.)
d
e
f
!
response
variable #
object #
conc.


Only Y and c known
Y
n
q
c
s´
d
z´
F
Only estimate of z is OK!


Input data Y
=
Analyte conc. c
×
Analyte spectrum s´ (est.)
a
b
c
response
variable #
object #
conc.
Residual data F1
=
Interfer. conc. d (est.)
×
Interfer. spectrum z´ (est.)
d
e
f
!


Analyte conc. c
Analyte spectrum s´
Interference conc. d
Y

vn
CA

# *Deshadowing via Informative Converse model:*
# Separating illumination / sample properties in HSI

### Data Model

In order to apply the described method, it is useful to first define a model for the measured data in each pixel:

$$\mathbf{Y} = \mathbf{C}\,\mathbf{S}^{\mathsf{T}} + \mathbf{D}\,\mathbf{Z}^{\mathsf{T}} + \mathbf{F}$$

With:

$$\begin{cases} \mathbf{Z} = \mathbf{A}^{\mathsf{T}}\mathbf{S}^{\mathsf{T}} + \mathbf{Z}^{\mathsf{T}}_{\perp\mathbf{S}} \\ \mathbf{C} = \mathbf{D}\mathbf{B} + \mathbf{C}_{\perp\mathbf{D}} \end{cases}$$

Where:

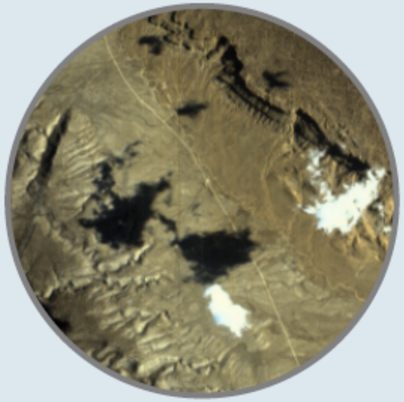
- $\mathbf{Y}$ is absorbance data obtained by sensor
- $\mathbf{C}\mathbf{S}^{\mathsf{T}}$ is the contribution of partially known effects (e.g. illumination variations, "shadows")
- $\mathbf{D}\mathbf{Z}^{\mathsf{T}}$ is the contribution of unknown effects (e.g. ground geology/biology variations)
- $\mathbf{F}$ is measurement noise, assumed normal
- $\mathbf{A}$ captures the non-orthogonality between $\mathbf{Z}$ and $\mathbf{S}$
- $\mathbf{B}$ captures the non-orthogonality between $\mathbf{C}$ and $\mathbf{D}$

### Assumptions

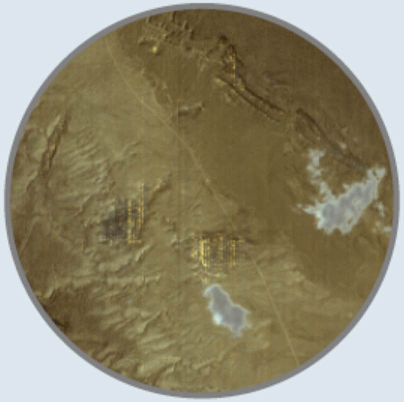
- Illumination effects are multiplicative in Reflectance
- $\mathbf{Y}$ data is given in Absorbance $(-\log_{10}(R))$
- Spectra of different illumination sources $\mathbf{S}$ are known

### Earth Observing-1

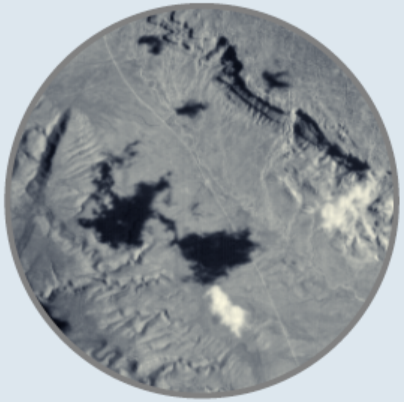
Data from the Hyperion instrument onboard the EO-1 Satellite. Data contains 200 bands in the VIS-NIR region.



Input data Y, in RGB



Deshadowed image, in RGB



"Shadow" (illumination change) image, $\widehat{CS}^{\mathsf{T}}$, in RGB

J.F. Fortuna and H. Martens, *J. Spectral Imaging* **6**, a2 (2017) Multivariate data modelling for de-shadowing of airborne hyperspectral imaging

# Fast decomposition of hyperspectral images

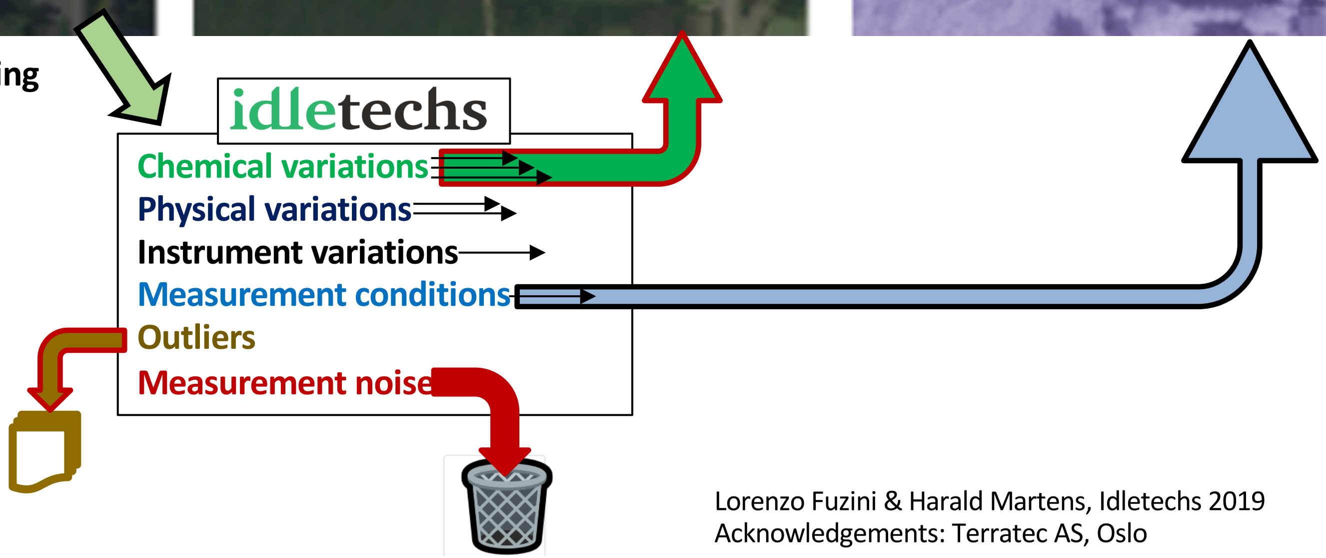
**Input HSI image, in RGB**

**Deshadowed image**

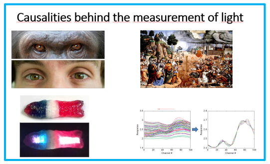
**Shadow image**

**Hybrid multivariate modelling of causalities in HSI:**

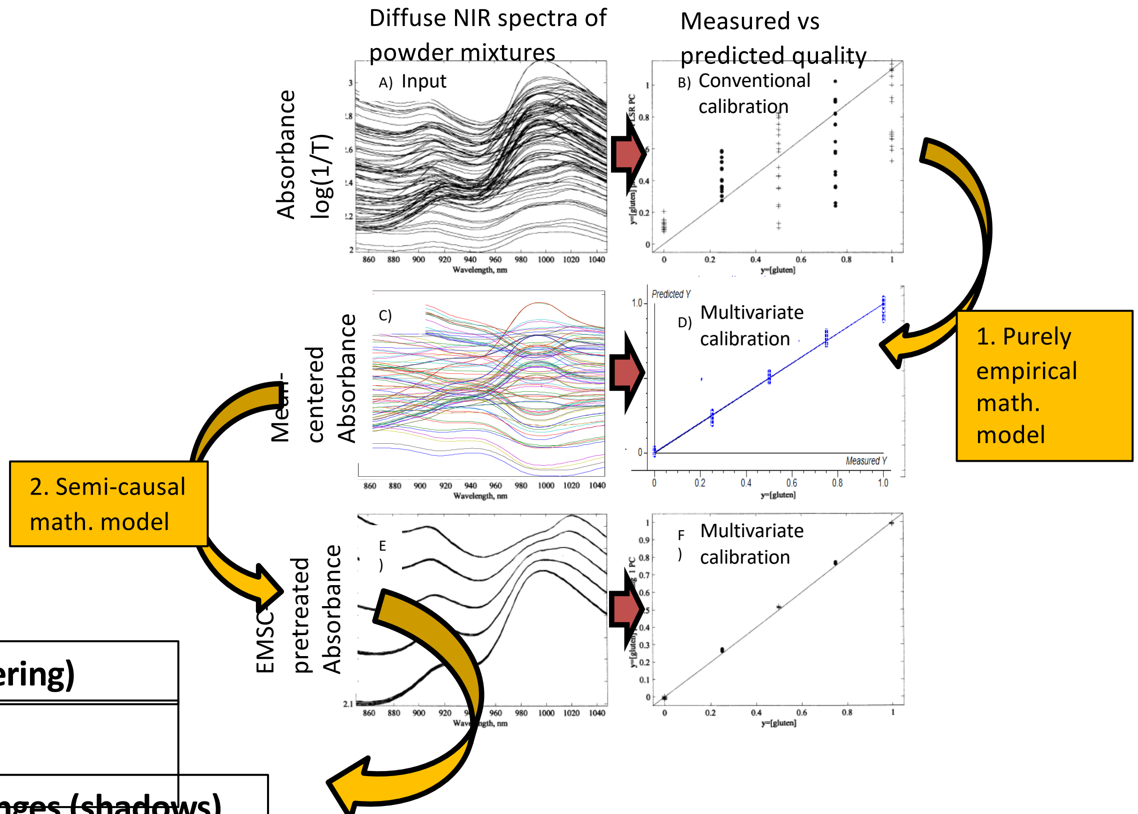
**Example from arial surveillance of biological resources by NEO HYSPEX camera**




Lorenzo Fuzini & Harald Martens, Idletechs 2019
Acknowledgements: Terratec AS, Oslo




# Non-additive multivariate modelling to separate light scattering and light absorbance




**2018 Explainable AI (XAI) based on chemometrics, e.g.:**

**Reflectance spectra**

- **Physics (scattering)**
- **Chemistry (absorbances)**
- **Condition changes (shadows)**
- **Sensor changes (Temp. effects)**

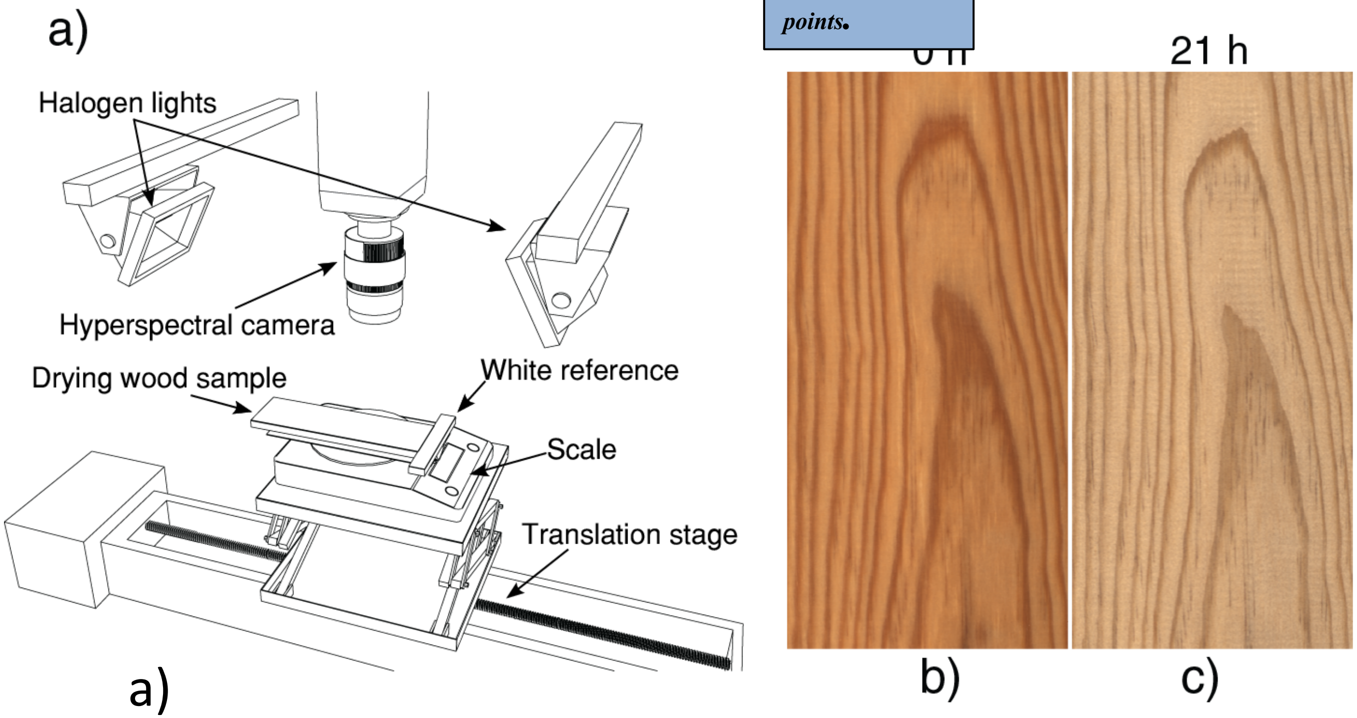
# Watch my hands now

**Hyperspectral video analysis: Hyperspectral image data streams interpreted by modeling known and unknown variations**

Authors: P. Stefansson[a], J. Fortuna[b, c], H. Rahmati[b], I. Burud[a], T. Konevskikh[a], H. Martens[b, c]

[a]Faculty of Science and Technology, Norwegian University of Life Sciences NMBU, Drøbakveien 31, 1430 Ås
[b]Idletechs AS, Havnegata 9, 7010 Trondheim Norway
[c]Department of Engineering Cybernetics, Norwegian University of Science and Technology NTNU, 7034 Trondheim Norway

*(2200 x 1070) pixels × 159 wavelengths × 150 time points.*




What are the causes that control wood drying ?

* Spectra?
* Spatial patterns?
* Time dynamics?

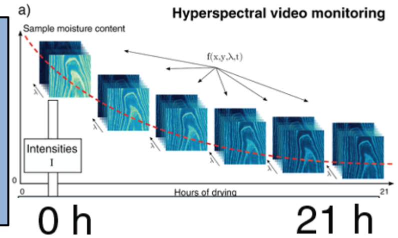
**Figure 2.12.1.** The experiment. a) Illustration of experimental setup used to measure the spectral reflectance and weight of a drying wood sample. b) RGB rendering of wood sample in wet state (drying time = 0 hours). c) RGB rendering of wood sample in dry state (drying time = 21 hours).

**Hyperspectral video analysis: Hyperspectral image data streams interpreted by modeling known and unknown variations**

Authors: P. Stefansson[a], J. Fortuna[b, c], H. Rahmati[b], I. Burud[a], T. Konevskikh[a], H. Martens[b, c]

[a]Faculty of Science and Technology, Norwegian University of Life Sciences NMBU, Drøbakveien 31, 1430 Ås
[b]Idletechs AS, Havnegata 9, 7010 Trondheim Norway
[c]Department of Engineering Cybernetics, Norwegian University of Science and Technology NTNU, 7034 Trondheim Norway

*(2200 x 1070) pixels × 159 wavelengths × 150 time points.*

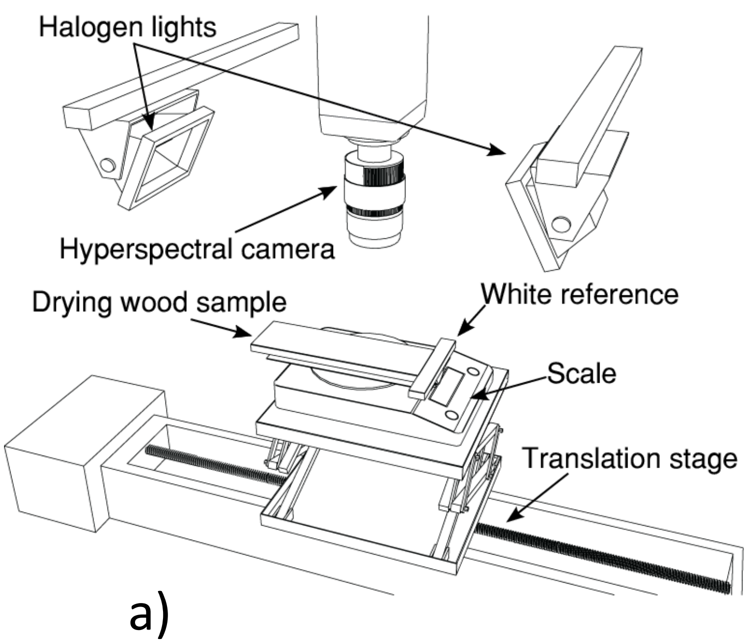



**>300 GB raw data from one experiment**

What are the causes that control wood drying ?

* Spectra?
* Spatial patterns?
* Time dynamics?

a)



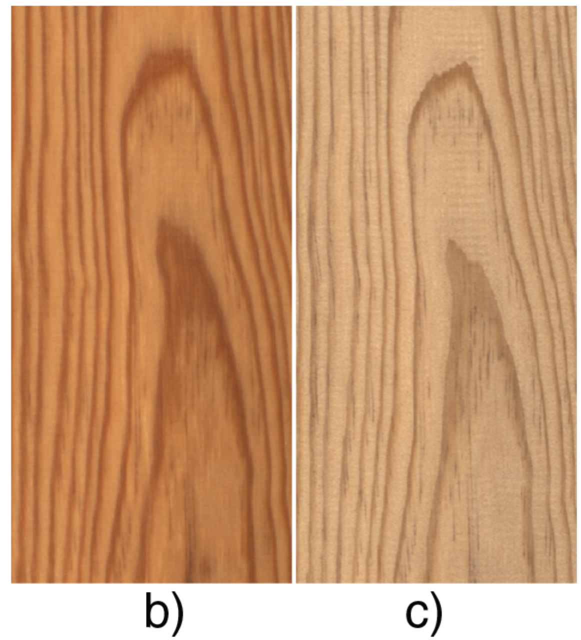
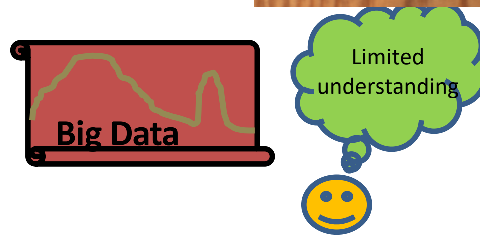




**Figure 2.12.1.** The experiment. a) Illustration of experimental setup used to measure the spectral reflectance and weight of a drying wood sample. b) RGB rendering of wood sample in wet state (drying time = 0 hours). c) RGB rendering of wood sample in dry state (drying time = 21 hours).

## Quantitative Big Data in Bio-sciences

>300 GB raw data from one experiment

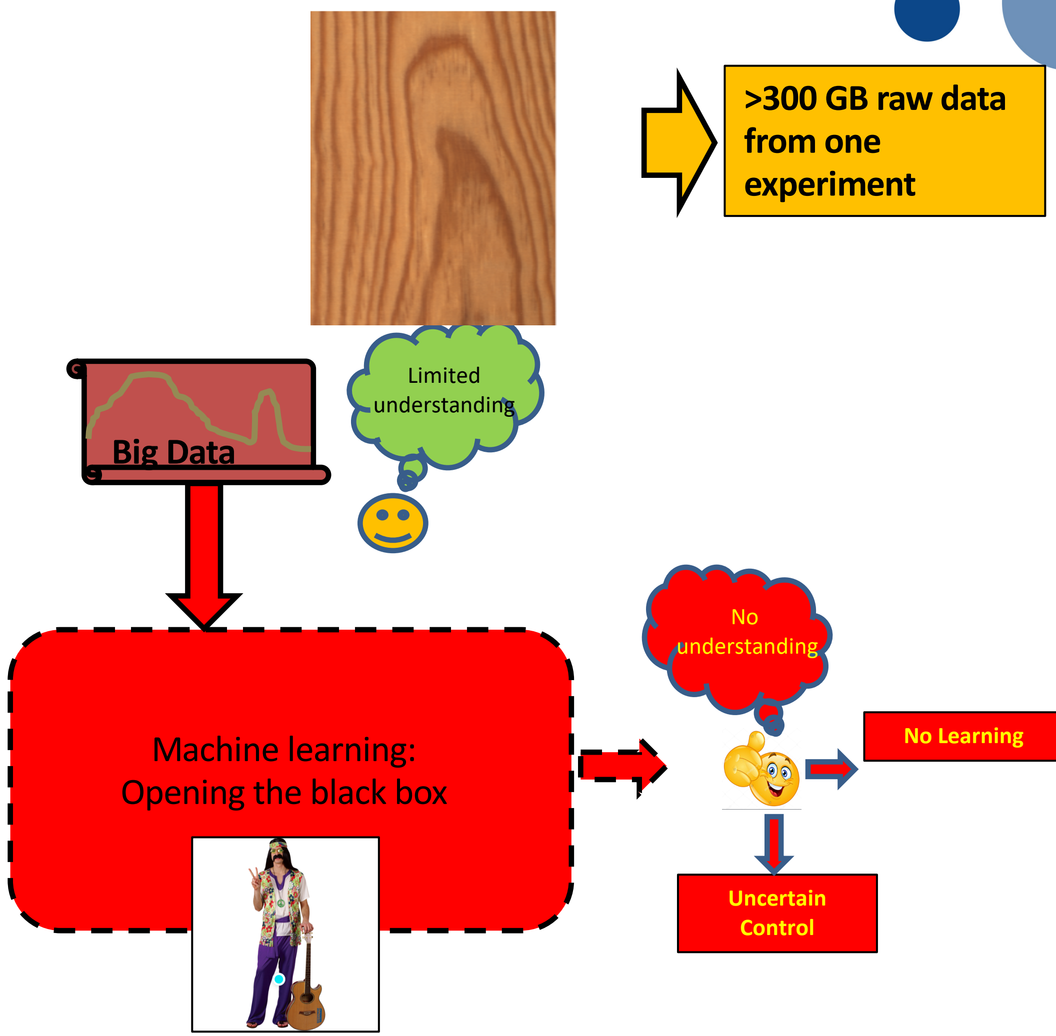



P. Stefansson, J. Fortuna, H. Rahmati, I. Burud, T. Konevskikh, H. Martens (2019): Hyperspectral video analysis: Hyperspectral image data streams interpreted by modeling known and unknown variations. *In press*.

idletechs

**Quantitative Big Data in Bio-sciences**

**« AI » :**

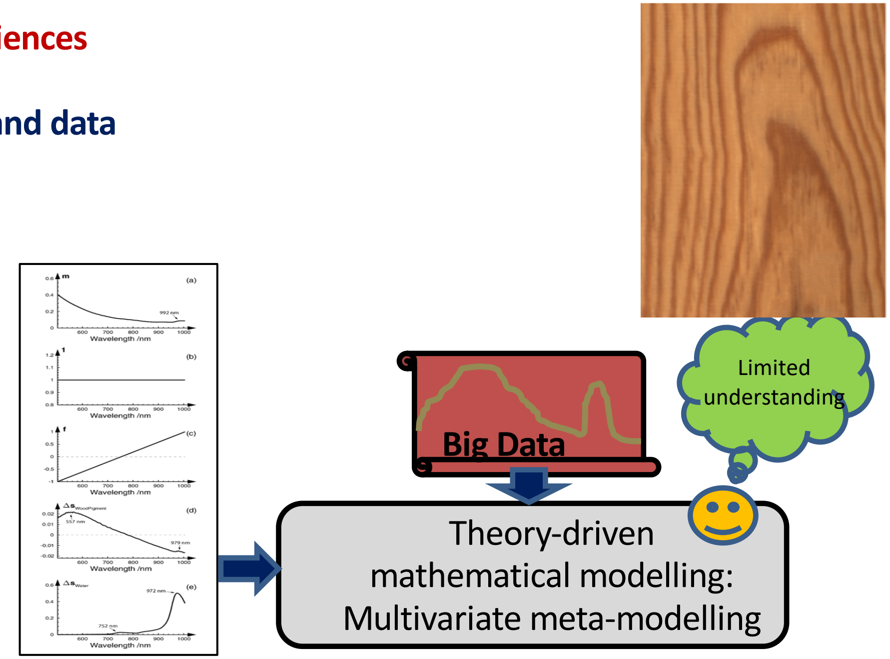



idletechs

## Quantitative Big Data in Bio-sciences

### « XAI » : Combine knowledge and data

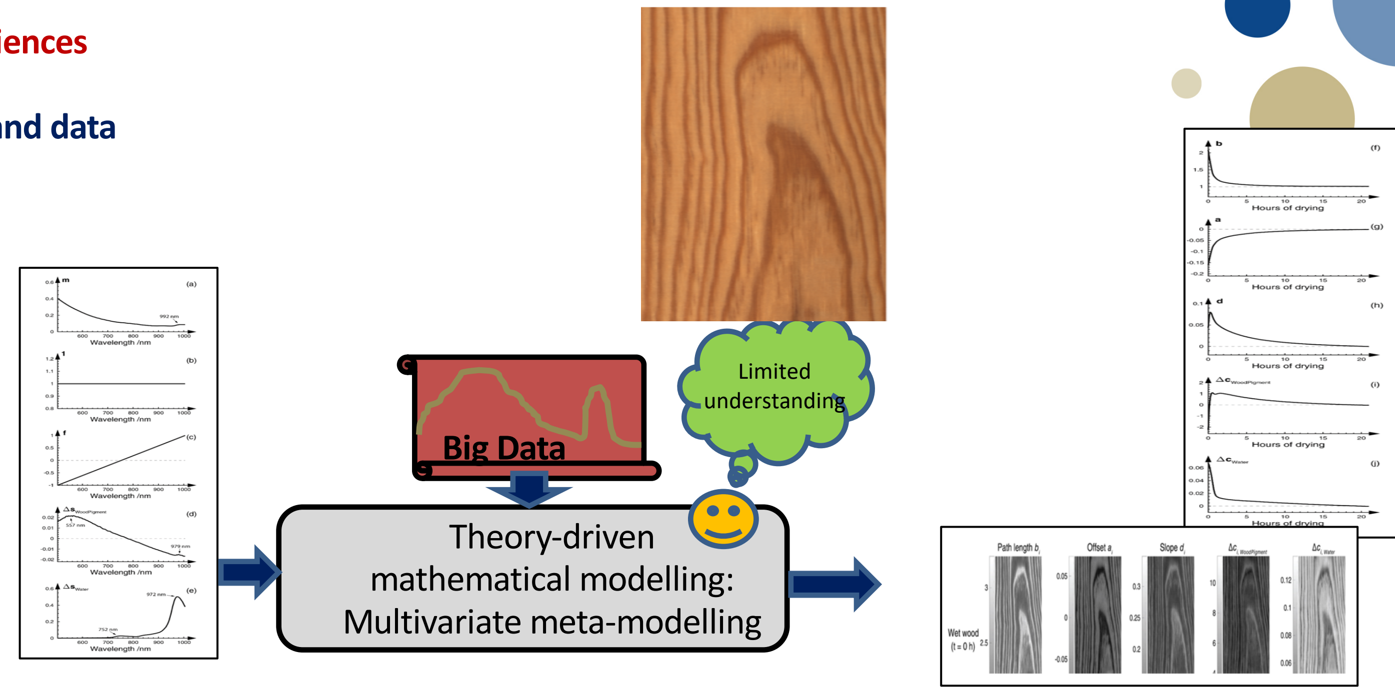
**>300 GB raw data from one experiment**




idletechs

## Quantitative Big Data in Bio-sciences

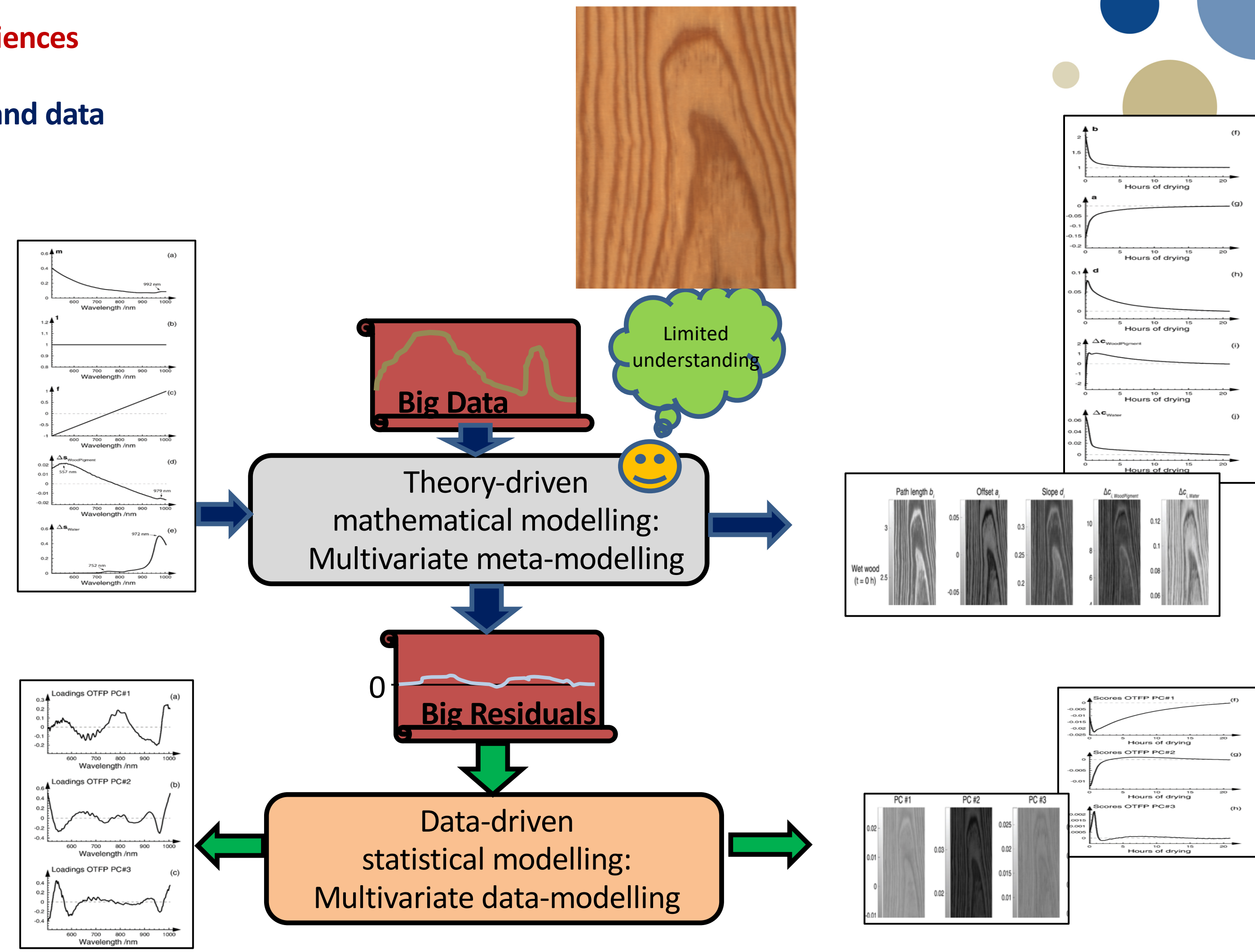
## « XAI » : Combine knowledge and data




idletechs

## Quantitative Big Data in Bio-sciences

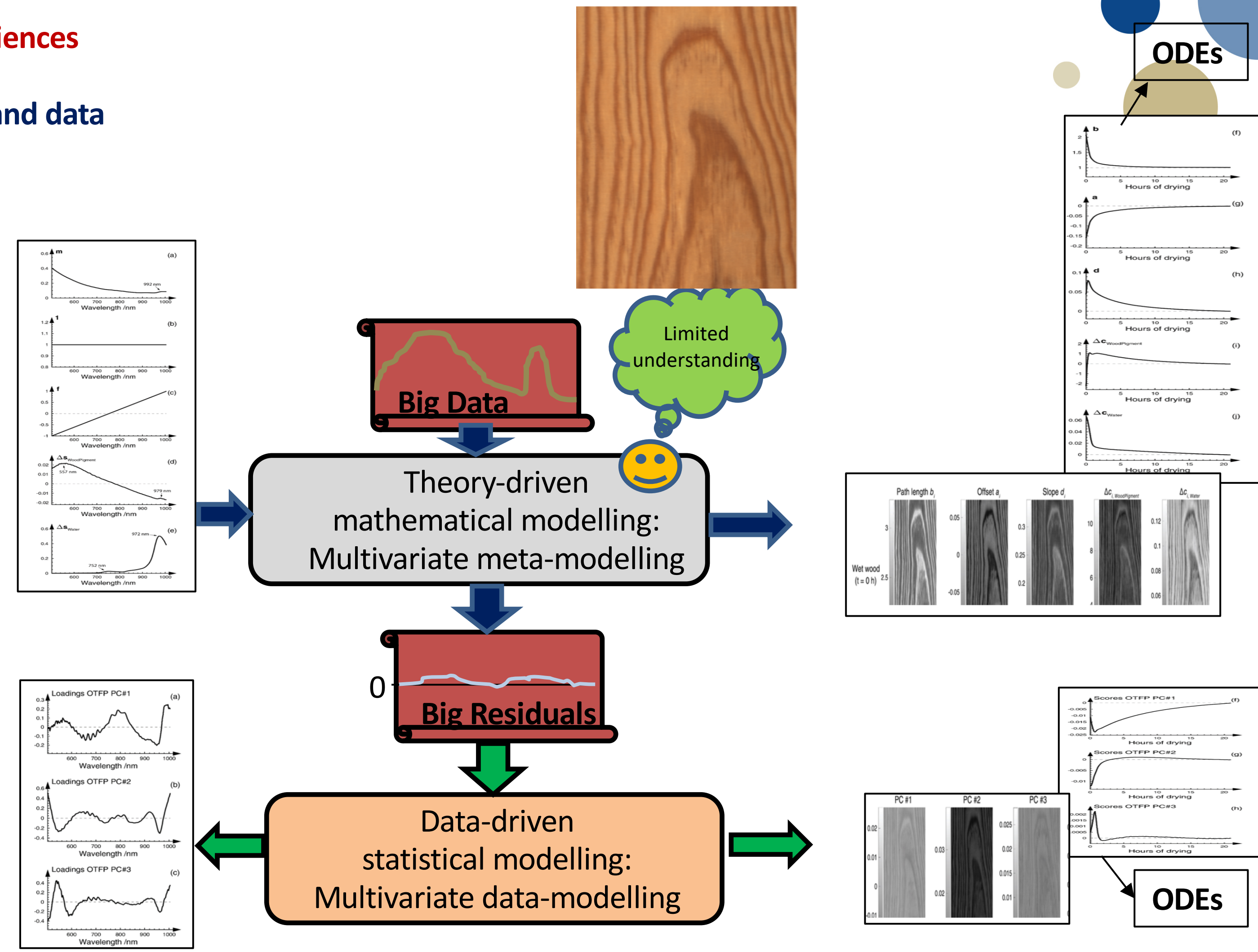
## « XAI » : Combine knowledge and data




idletechs

## Quantitative Big Data in Bio-sciences

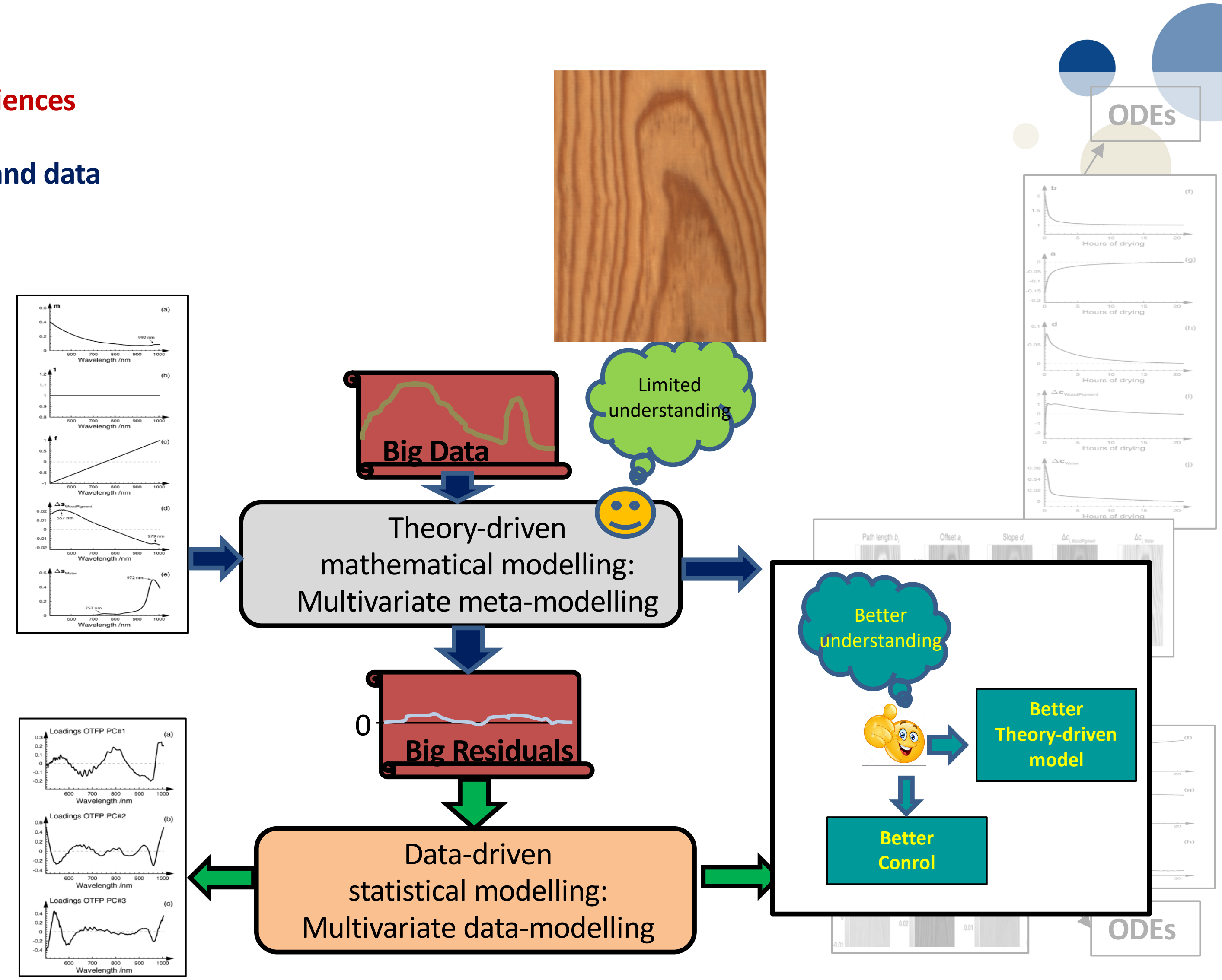
## « XAI » : Combine knowledge and data




idletechs

## Quantitative Big Data in Bio-sciences

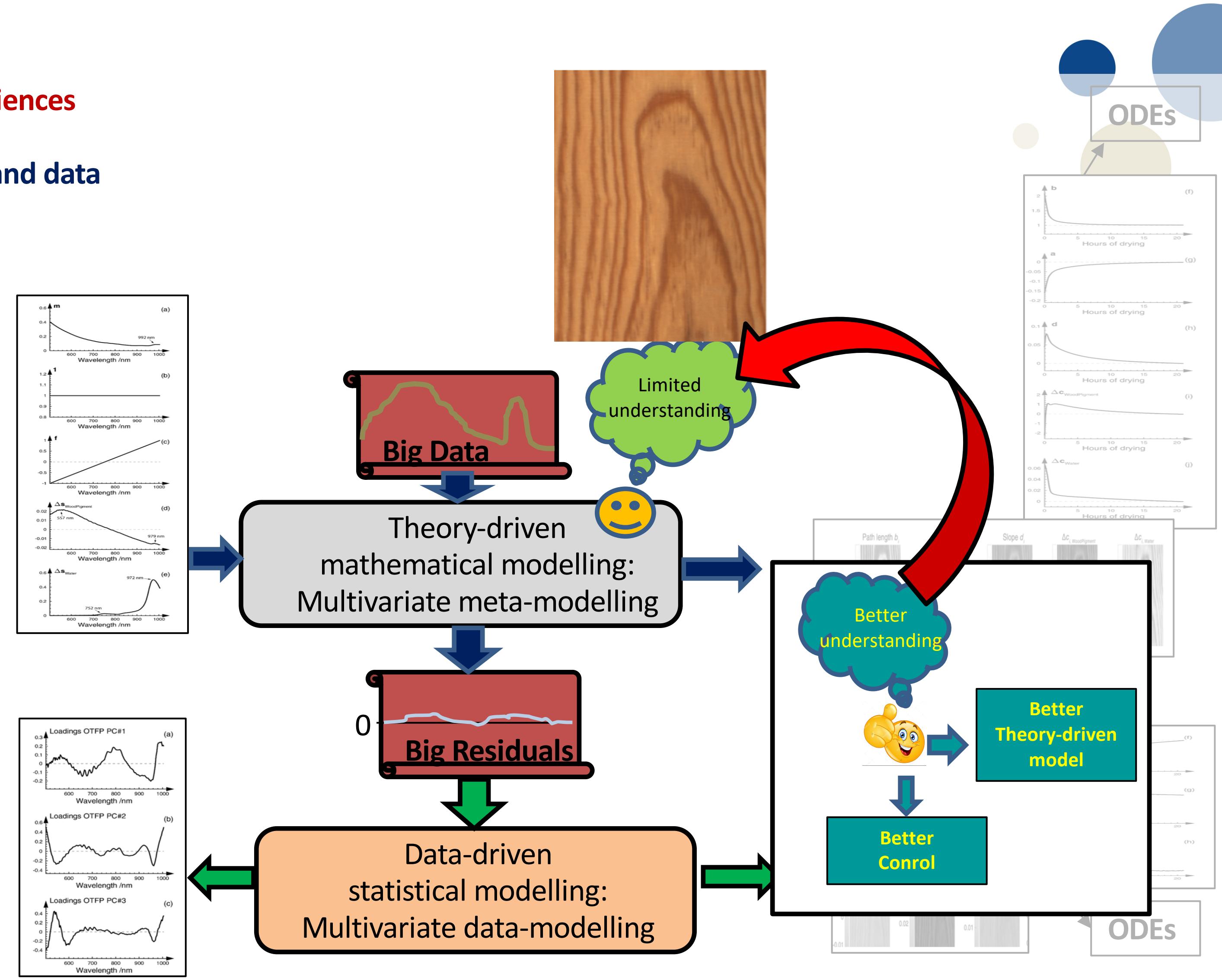
## « XAI » : Combine knowledge and data




idletechs

## Quantitative Big Data in Bio-sciences

## « XAI » : Combine knowledge and data




idletechs

# What is Design of Experiments?

Design of Experiments (DoE) is the pre-planned, **systematic variation** of **controllable** experimental factors that **induce a response** in a system.

The factors are measured in such a way that the **minimum effort** is required to gain a **maximum amount of information**.

# What is Design of Experiments?

Design of Experiments (DoE) is the pre-planned, **systematic variation** of **controllable** experimental factors that **induce a response** in a system.

The factors are measured in such a way that the **minimum effort** is required to gain a **maximum amount of information**.

Two uses:

1) Phyiscal experiments in lab or industry
2) Computer experiments to study the behavior of a mathematical model, and to speed it up.

# Summary (1/2)

- DoE is the best way of generating meaningful experiments that will provide the **maximum information** in the **minimal experimental effort**
- Designed experiments can be performed sequentially, i.e. **more information can be added** if need be, to an existing design
- Fractional factorials can be **simplified** when factors are found to be insignificant, resulting in more precise results, without any further experimentation being done
- Many factors can be analyzed in a **small number of experiments** to screen out important factors

# Summary (2/2)

- Remember to check the **residuals** of the model to see if the assumptions of the ANOVA hold or to detect outliers
- When a small number of factors have been isolated, the design can be **extended** to become an optimization design
- When dealing with mixtures, the experimental designs become **constrained**. A special class of designs must be considered to analyze these.

# Designed computer simulations to find the behavioural repertoire of a complicated model

**Example: Speed-up of a mathematical Finite-element model of facial muscles (Tim Wu, Auckland Bio-engineering Institute, NZ)**

**18 parameters, 4 levels of each:**
**> $10^{10}$ combinations?**

CPU time for each new simulation, using the original model:
**>2 hours**

128 simulations, then PLSR-based multivariate metamodel:
CPU time for each n ew simulation, using its multivariate metamodel:
**< 0.01 second**

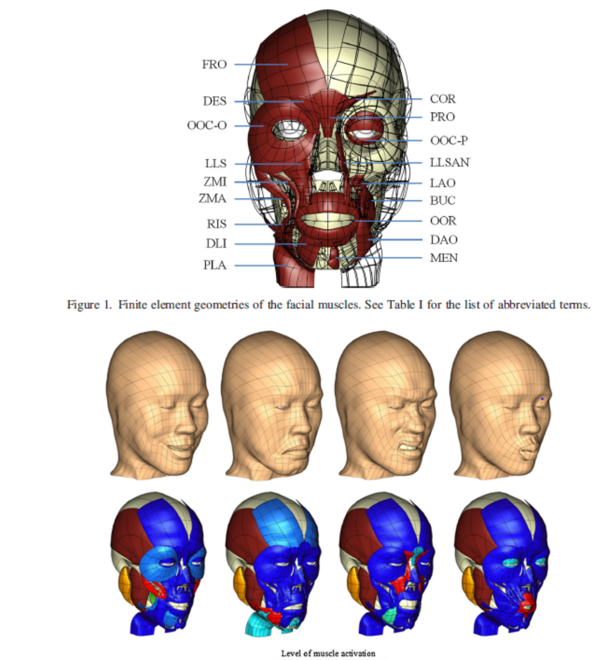

TOWARDS FAST AND REALISTIC GENERATION OF FACIAL EXPRESSIONS



Figure 1. Finite element geometries of the facial muscles. See Table I for the list of abbreviated terms.



Figure 2. Biomechanical simulations of the expressions of (from left to right) joy, sadness, snarl and the kissing gesture, by activating the corresponding facial muscles.


INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN BIOMEDICAL ENGINEERING
*Int. J. Numer. Meth. Biomed. Engng.* (2014)
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/cnm.2646


Emulating facial biomechanics using multivariate partial least squares surrogate models


Tim Wu[1,*,†], Harald Martens[2], Peter Hunter[1] and Kumar Mithraratne[1]

# *Designed computer simulations to find the behavioural repertoire of a complicated model*

**Example: Speed-up of a mathematical Finite-element model of facial muscles (Tim Wu, Auckland Bio-engineering Institute, NZ)**

**18 parameters, 4 levels of each:**
**> $10^{10}$ combinations?**

CPU time for each new simulation, using the original model:
**>2 hours**

TOWARDS FAST AND REALISTIC GENERATION OF FACIAL EXPRESSIONS

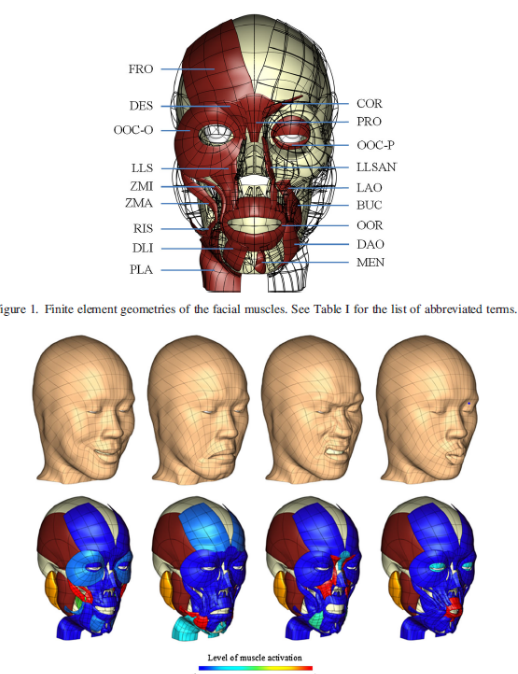



Figure 1. Finite element geometries of the facial muscles. See Table I for the list of abbreviated terms.

Figure 2. Biomechanical simulations of the expressions of (from left to right) joy, sadness, snarl and the kissing gesture, by activating the corresponding facial muscles.

INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN BIOMEDICAL ENGINEERING
*Int. J. Numer. Meth. Biomed. Engng.* (2014)
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/cnm.2646

Emulating facial biomechanics using multivariate partial least squares surrogate models

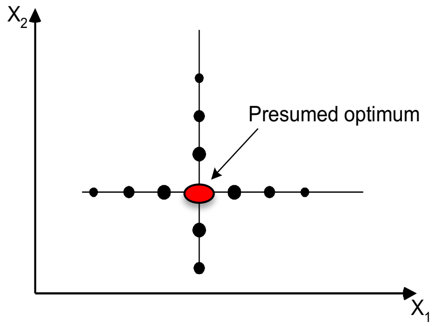
Tim Wu[1,*,†], Harald Martens[2], Peter Hunter[1] and Kumar Mithraratne[1]

# One variable at the time (OVAT)

In order to establish a relationship between cause and effect,
each cause must be investigated separately,
all other conditions being fixed.

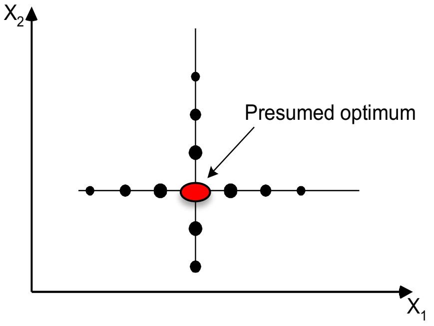
# How to span the experimental space

The Classical Approach:
(OVAT)

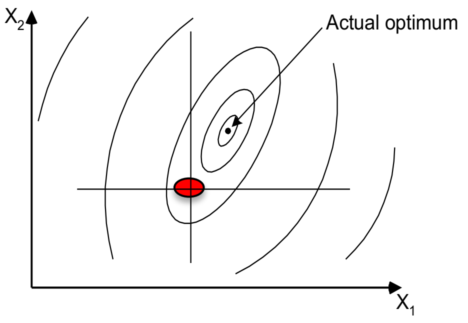


# How to span the experimental space

The Classical Approach:
(OVAT)

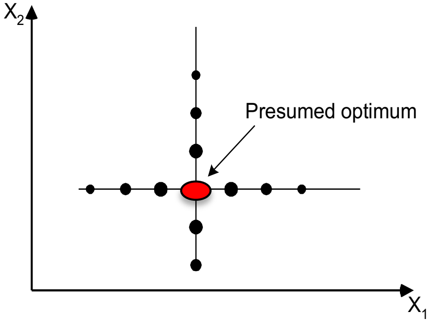



What can go wrong?



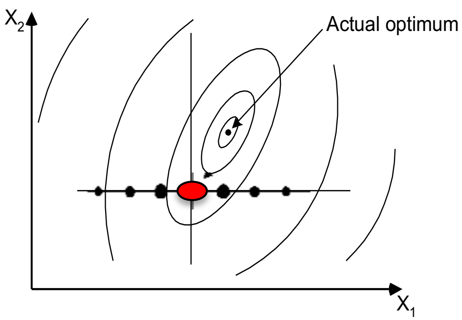
# How to span the experimental space

The Classical Approach:
(OVAT)



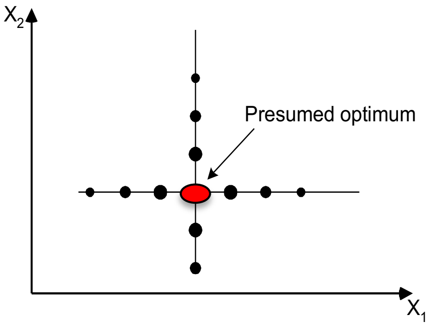

What can go wrong?

*Optimality along a ridge*



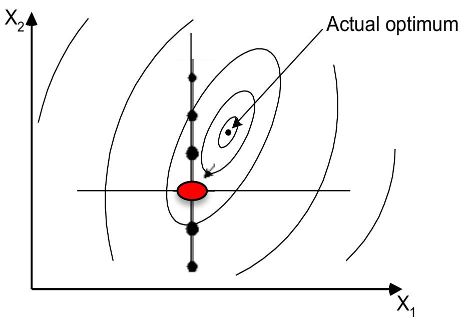
# How to span the experimental space

The Classical Approach:
(OVAT)



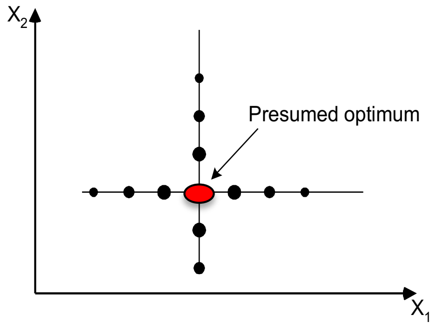

What can go wrong?

*Optimality along a ridge*

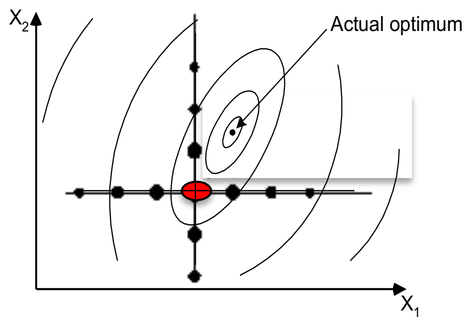


# How to span the experimental space

The Classical Approach:
(OVAT)



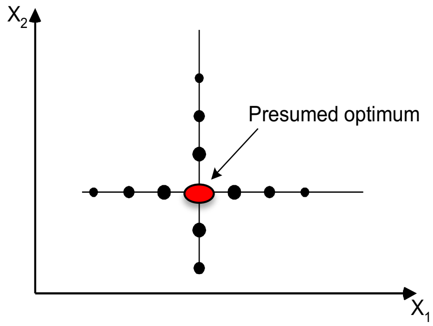

What can go wrong?




*Optimality along a ridge*
***Found solution looks OK, but is far from the optimum***

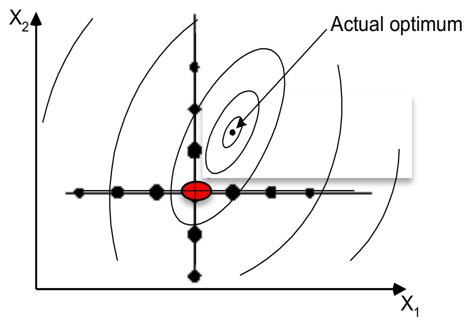
# How to span the experimental space

The Classical Approach:
(OVAT)



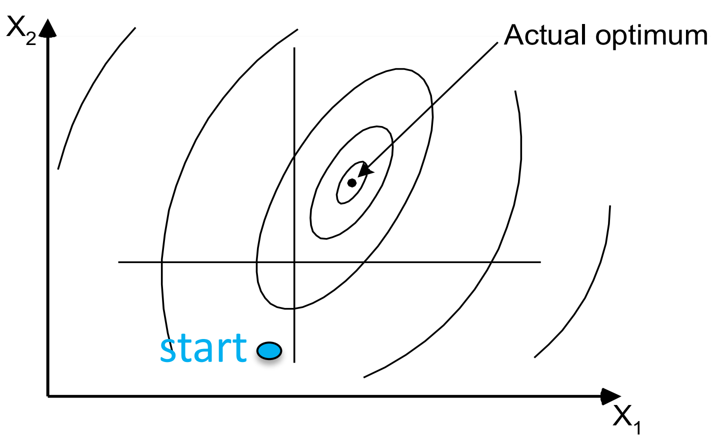

What can go wrong?

*Optimality along a ridge*



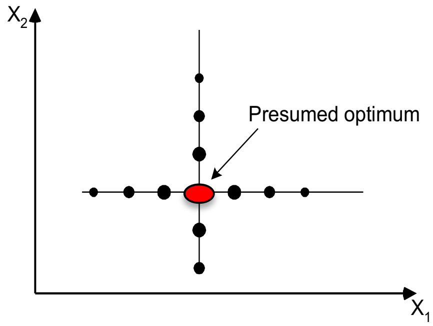

How can we do it better?

Iterative, empirical process:
One experiment at a time,
SIMPLEX algorithm to plan next experiment



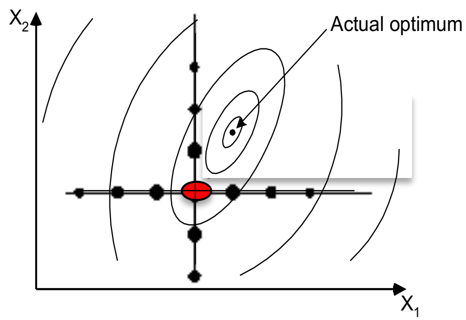
# How to span the experimental space

The Classical Approach:
(OVAT)



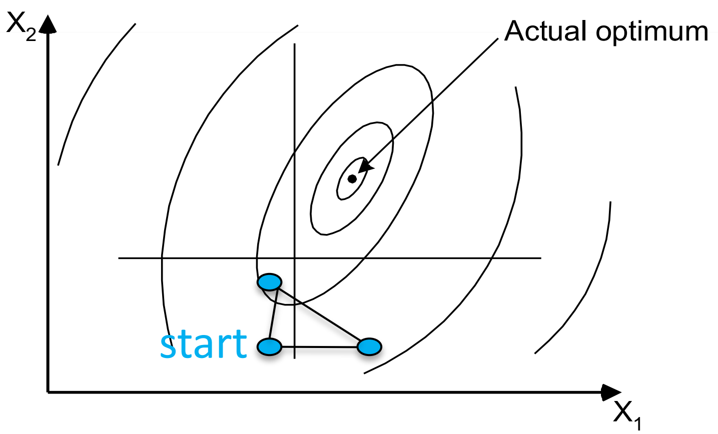

What can go wrong?

*Optimality along a ridge*



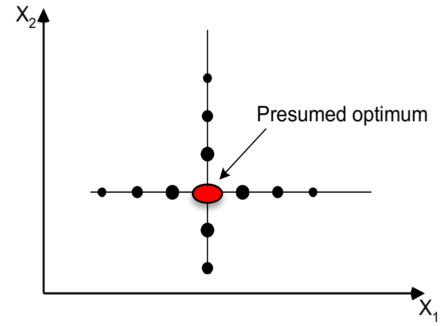

How can we do it better?

Iterative, empirical process:
One experiment at a time,
SIMPLEX algorithm to plan next experiment



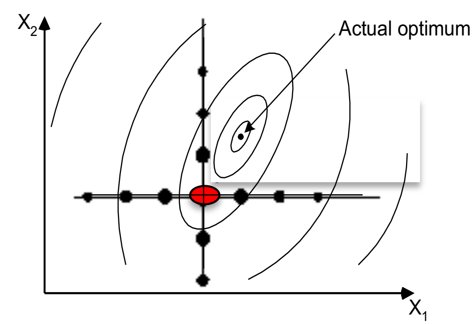
# How to span the experimental space

The Classical Approach:
(OVAT)

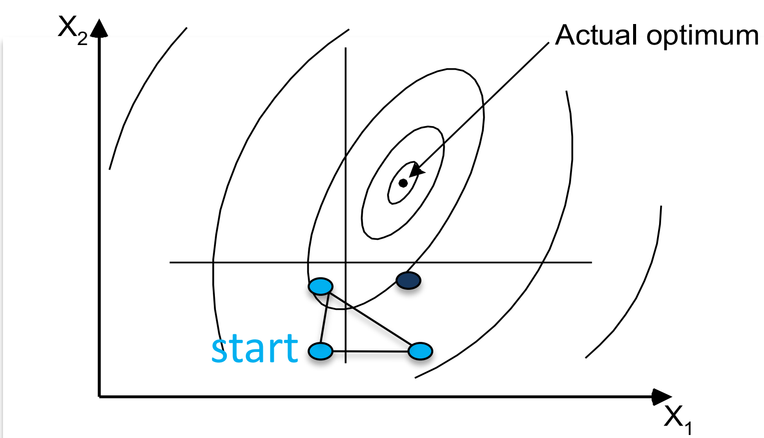



What can go wrong?

*Optimality along a ridge*



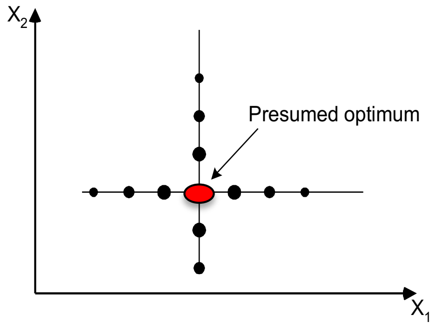

How can we do it better?




Iterative, empirical process:
One experiment at a time,
SIMPLEX algorithm to plan next experiment

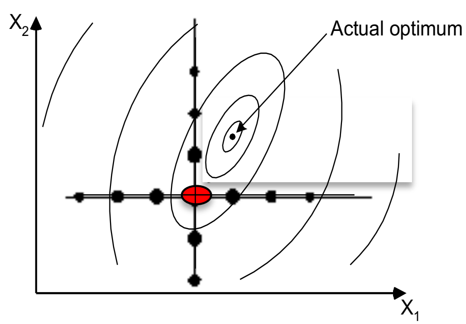
# How to span the experimental space

The Classical Approach:
(OVAT)



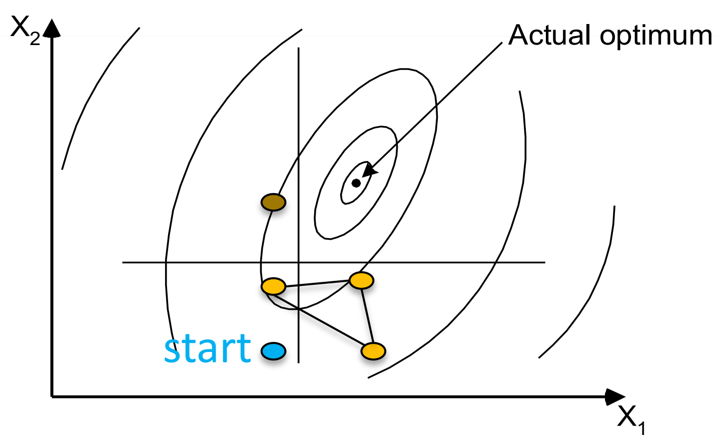

What can go wrong?

*Optimality along a ridge*



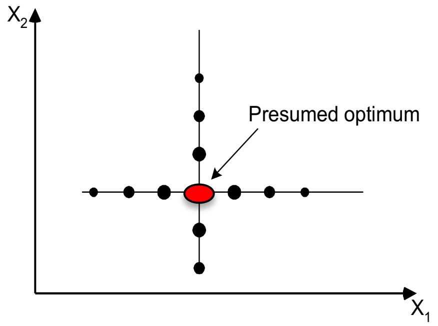

How can we do it better?

Iterative, empirical process:
One experiment at a time,
SIMPLEX algorithm to plan next experiment



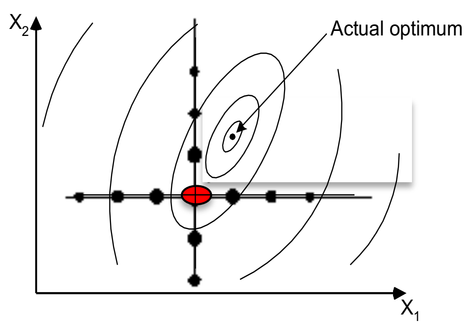
# How to span the experimental space

The Classical Approach:
(OVAT)



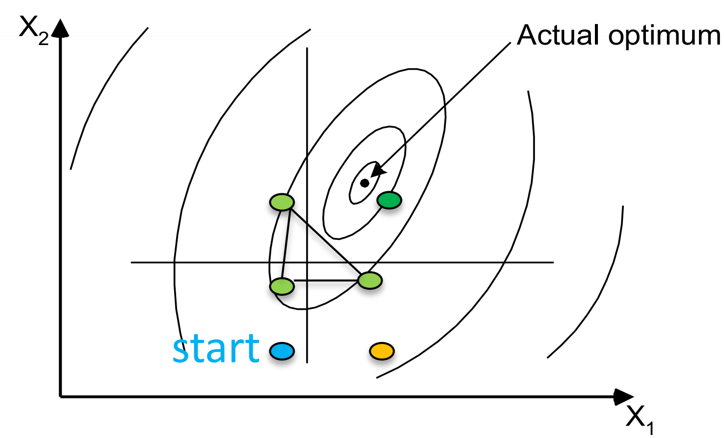

What can go wrong?

*Optimality along a ridge*



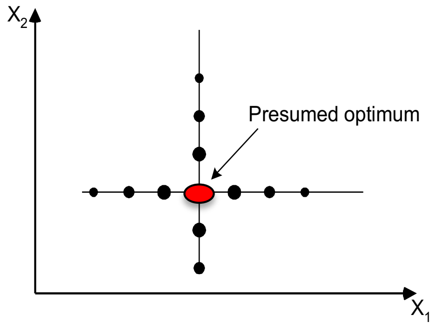

How can we do it better?

Iterative, empirical process:
One experiment at a time,
SIMPLEX algorithm to plan next experiment



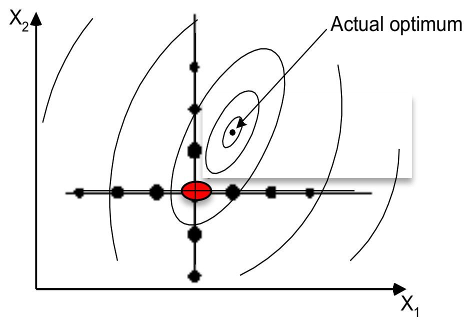
# How to span the experimental space

The Classical Approach:
(OVAT)



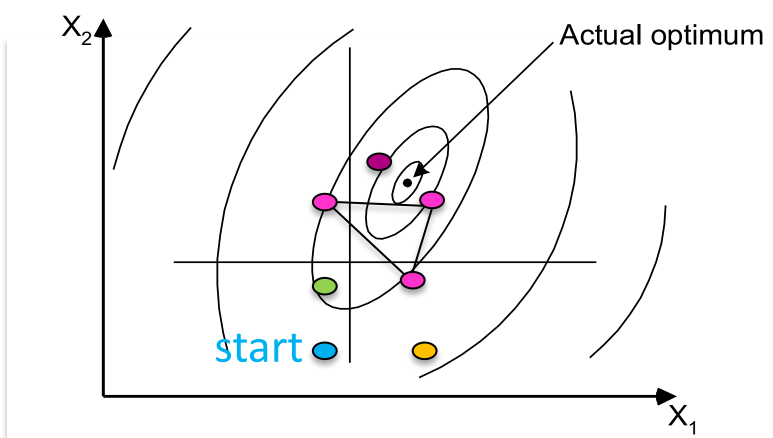

What can go wrong?

*Optimality along a ridge*



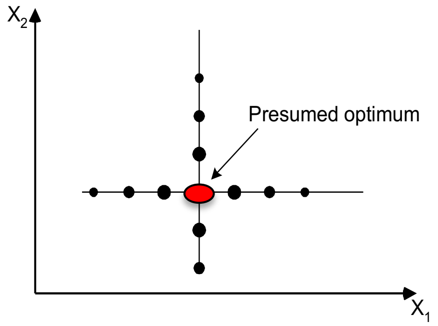

How can we do it better?

Iterative, empirical process:
One experiment at a time,
SIMPLEX algorithm to plan next experiment



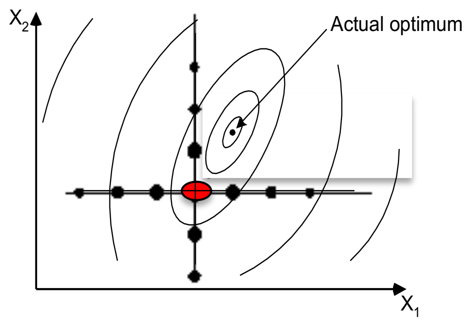
# How to span the experimental space

The Classical Approach:
(OVAT)



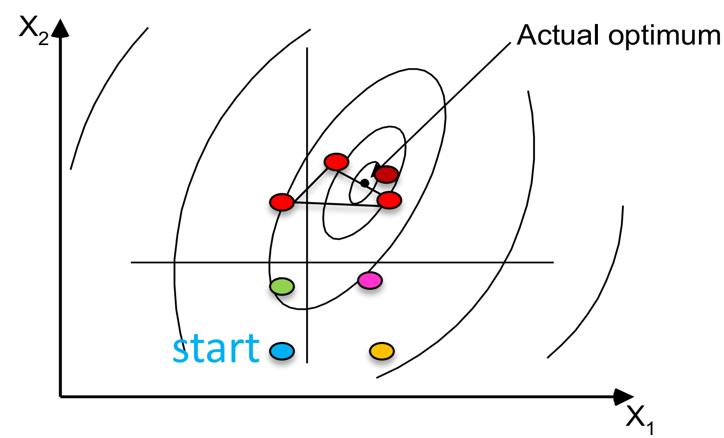

What can go wrong?

*Optimality along a ridge*



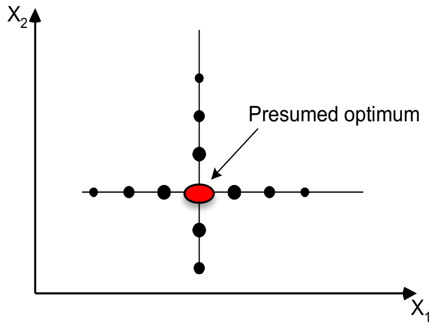

How can we do it better?

Iterative, empirical process:
One experiment at a time,
SIMPLEX algorithm to plan next experiment



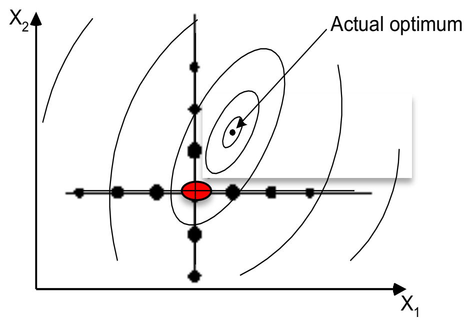
# How to span the experimental space

The Classical Approach:
(OVAT)



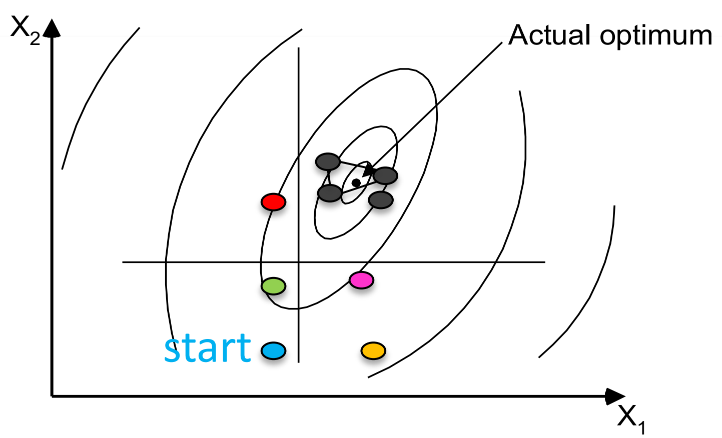

What can go wrong?

*Optimality along a ridge*



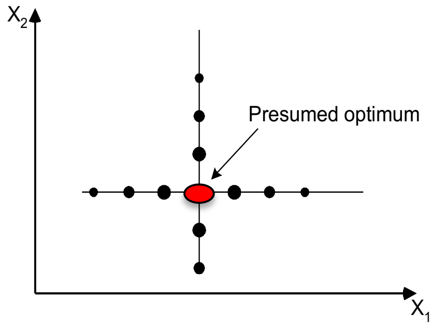

How can we do it better?

Iterative, empirical process:
One experiment at a time,
SIMPLEX algorithm to plan next experiment

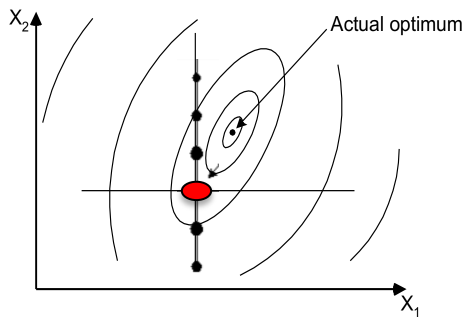


# OLD How to span the experimental space

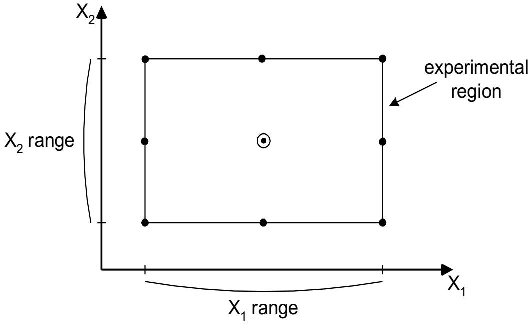
The Classical Approach:
(OVAT)




What can go wrong?




How can we do it better?




Systematic experimental process:
Several experiments at a time,according to a statistical design.
Analysis of results: ANOVA, PLS Regression etc

# Summary (1/2)

- DoE is the best way of generating meaningful experiments that will provide the **maximum information** in the **minimal experimental effort**
- Designed experiments can be performed sequentially, i.e. **more information can be added** if need be, to an existing design
- Fractional factorials can be **simplified** when factors are found to be insignificant, resulting in more precise results, without any further experimentation being done
- Many factors can be analyzed in a **small number of experiments** to screen out important factors

# Advantages of Experimental Design

| | OVAT | Experimental Design |
|---|---|---|
| Main effects | Not estimated ? | Estimated |
| Interactions | Not detected | Detected and estimated |
| Experimental Variability | 100% impact | Reduced impact |
| Number of experiments | Unknown | Known per step |
| Best solution<br>If no solution | ???<br>??? | Spotted<br>Detected |
| Several responses | Difficult | As easy as 1 response |
| New objectives | Start all over again! | Re-use existing results |

# Main types of designs

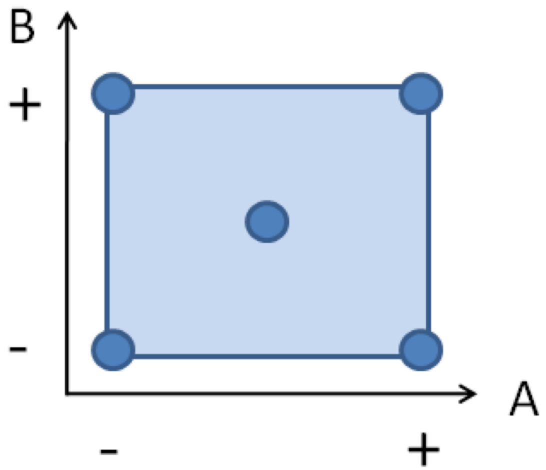
| Type of design | Objective |
|---|---|
| • Fractional factorial | Find main effects |
| • Full factorial | Find main effects and interactions |
| • Optimization designs | Find optimal settings for a response surface |
| • Mixture designs | Find the optimal recipe of a mixture |
| • D-optimal designs | Designs with constraints |

# The full factorial design

- Motivation for use:
  - Simplest design situation
  - Basis for many other designs
  - Optimal for detecting main effects and their interactions

# 2-level full factorial designs

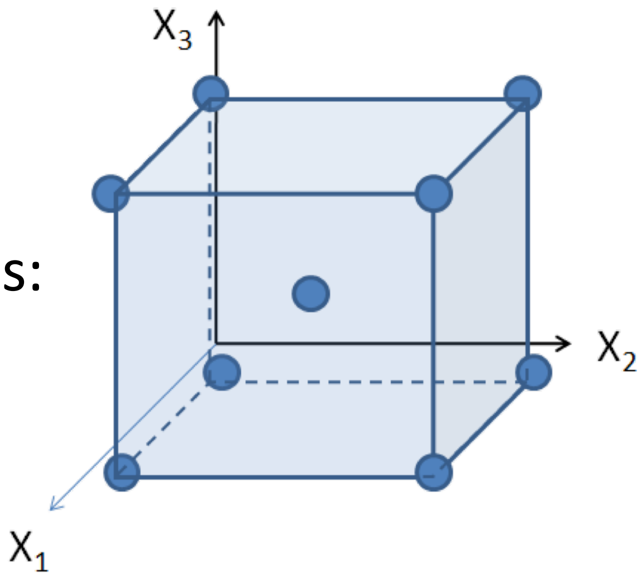
2 X-variables:




| run # | $X_1$ | $X_2$ |
|---|---|---|
| 2 | - | - |
| 4 | - | + |
| 6 | + | - |
| 1 | + | + |
| 3 | 0 | 0 |
| 5 | 0 | 0 |

$2^2$ experiments (+ centre samples)

3 X-variables:

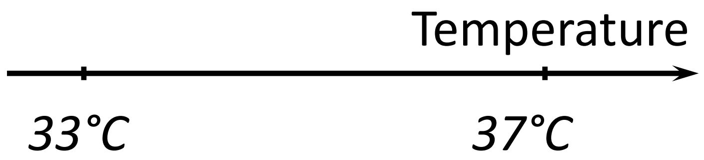


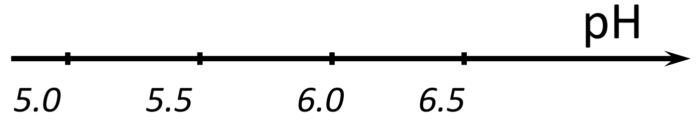

| run # | $X_1$ | $X_2$ | $X_3$ |
|---|---|---|---|
| 2 | - | - | - |
| 11 | - | - | + |
| 5 | - | + | - |
| 8 | - | + | + |
| 4 | + | - | - |
| 1 | + | - | + |
| 9 | + | + | - |
| 7 | + | + | + |
| 3 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 |
| 10 | 0 | 0 | 0 |

$2^3$ experiments (+ centre samples)

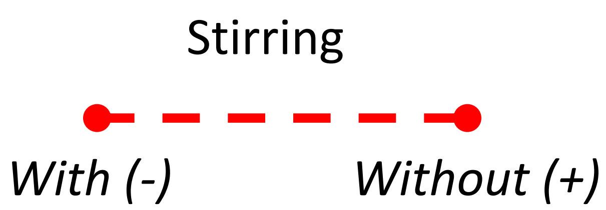
# Levels of Design Variables

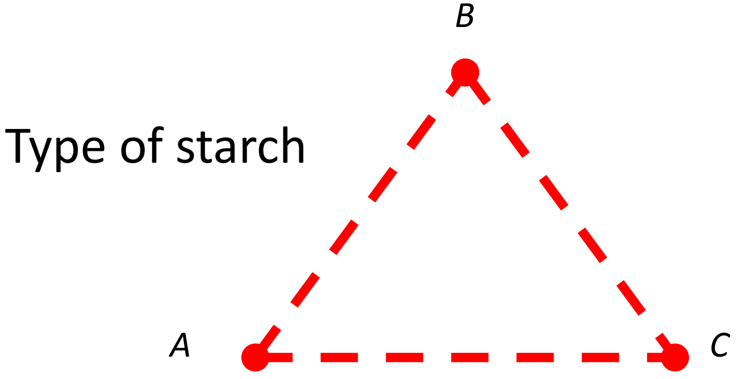
- Continuous variables
  - Range: Low to high
  - List of values




- Category variables



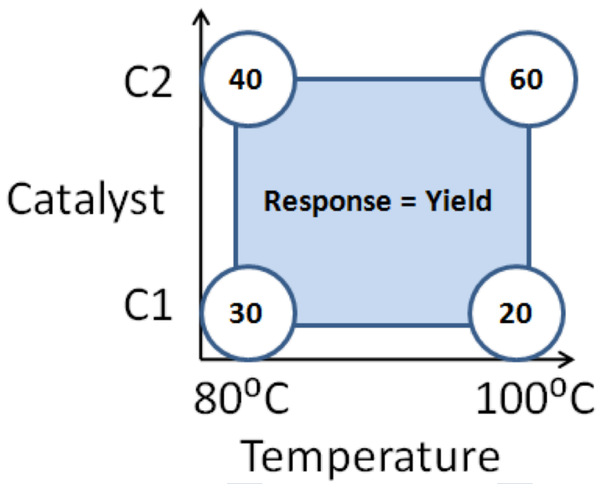
# Additional experiments

- Center samples
  - To detect curvature
  - To estimate error variance
  - Category?
    - One center point at each level

- Replicated samples
  - Replication of the factorial points
  - More precise estimates of error variance

# Main Effects

A simple experimental design:




Average of T at Low Level = (30+40)/2 =**35**

Average of C at High Level = (20+60)/2 =**40**

**Main Effect of Temperature =40-35 = +5**

**Main effect**: Temperature on Yield

**Main effect**: Catalyst on Yield

Average of C at High Level = (40+60)/2 =**50**

Average of C at Low Level = (30+20)/2 =**25**

**Main Effect of Catalyst**

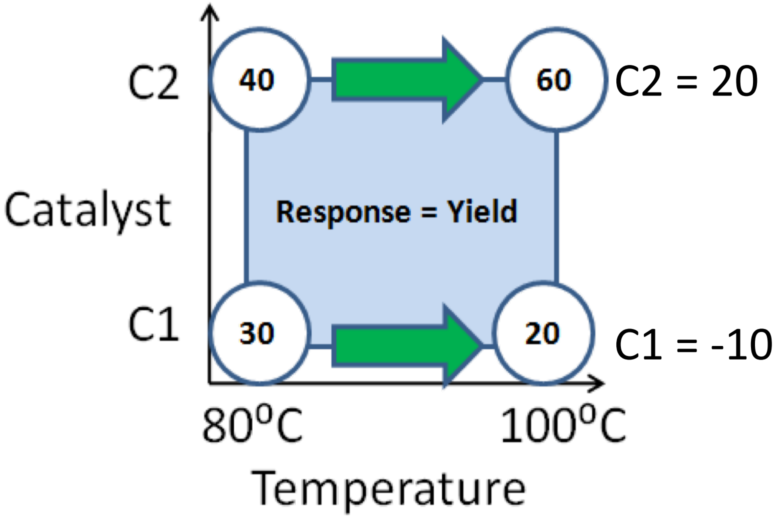
**=50-25**

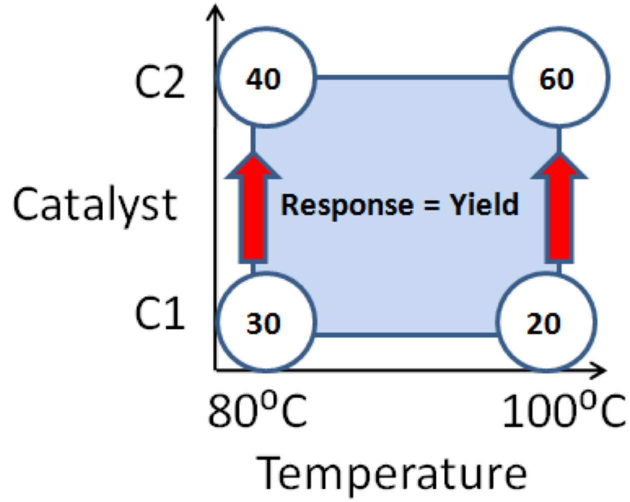
**= +25**

The Main Effects indicate what the overall effect a single variable has on the overall response

# Interactions

## **Interaction:** Catalyst*Temperature




**Effect of Temperature on Catalyst**

C2 = 60 – 40 =20
C1 = 20 – 30 = -10

Interaction = (20 – (-10))/2 **= +15**

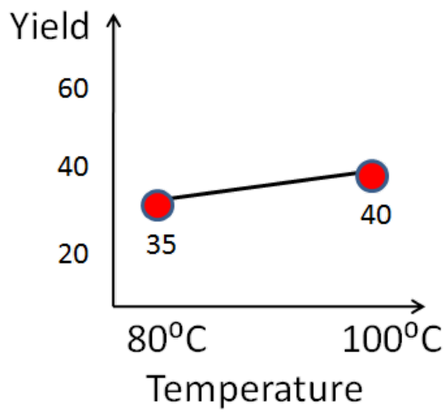
**Effect of Catalyst on Temperature**

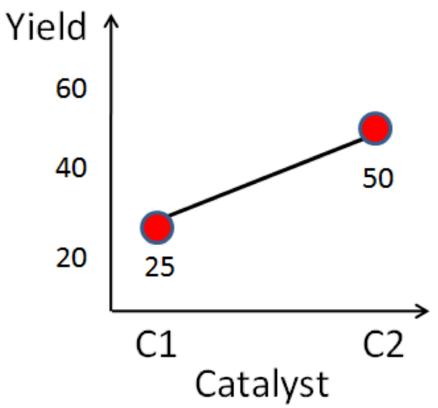
T2 = 60 – 20 =40
C1 = 40 – 30 =10

Interaction = (40 – 10)/2 **= +15**

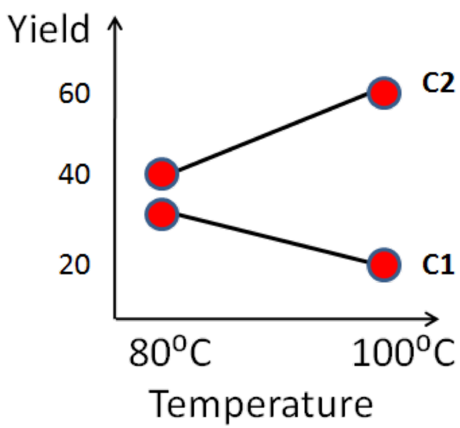
# Interpreting Effects

Main effects




Interactions

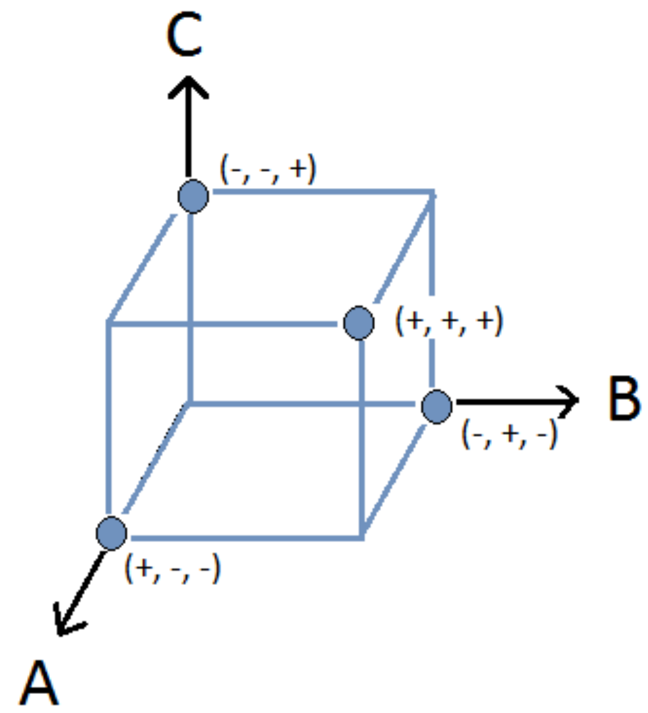


# Fractional factorial designs

- Full factorial expensive if many variables
- Often higher order interactions can be neglected → Fractional factorial design
- Subset of the full factorial design
- Experiments are systematically chosen to cover the widest possible design space
- Introduces confounding between the model terms -> not all effects can be estimated independently of other terms

# 2-level Fractional factorial design

- 3 design variables A, B, C:
- $2^{3-1}$ design, C = AB
- All main effects estimated in $2^{3-1}$ = 4 runs
- Main effect C confounded with interaction AB

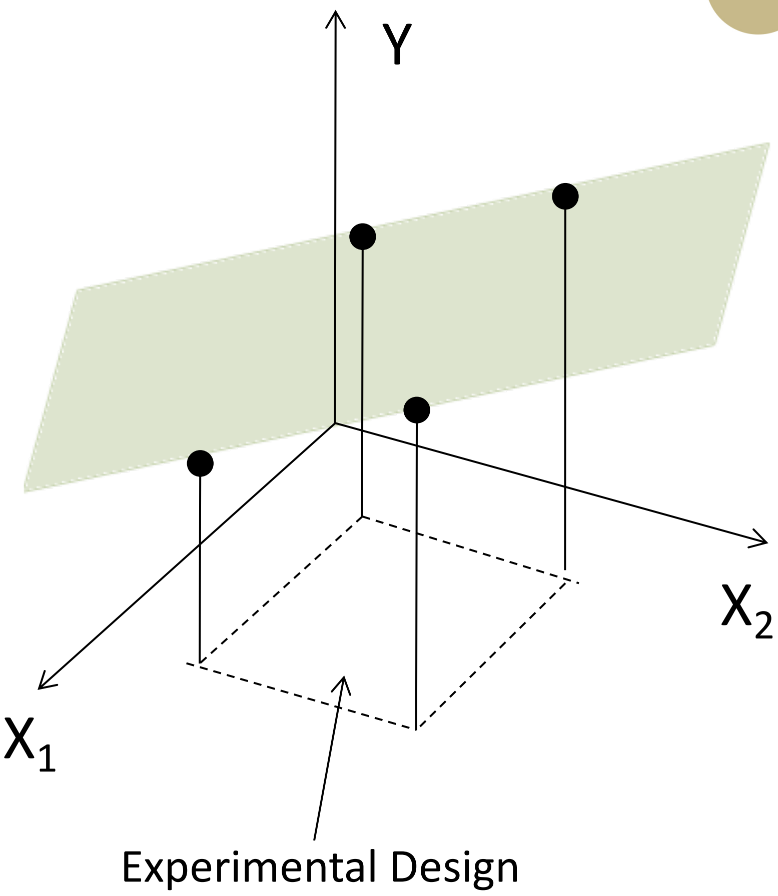


# Regression methods: MLR

MLR – Multiple Linear Regression

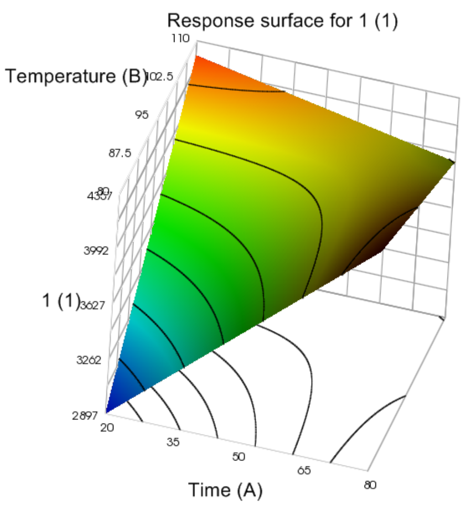
- The criterion of closeness between observed and fitted values is called the Least Squares criterion : the plane should lie where it minimizes the sum of squares of all residuals.

- The method of choice for orthogonal experimental designs
- The ANOVA table is calculated from the regressions coefficients in most implementations; in the earlier days by means of square sums (programming it yourself should take 10-15 minutes ☺)



# Response Surface: Overview

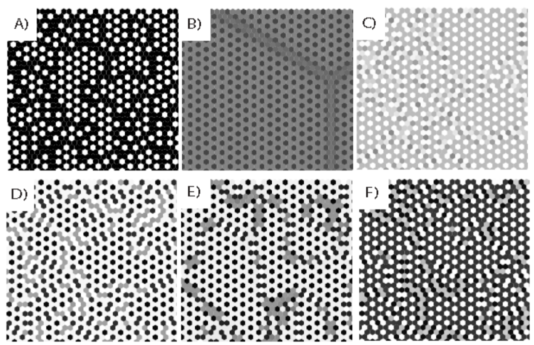
- Main results
  - Detailed ANOVA table
  - Normal probability of residuals
  - Contour/landscape plot
- Additional plots
  - Residuals versus predicted Y (various options)
  - Regression coefficients
  - Predicted versus measured Y



# Design of computer experiments

# Bruk sansene i naturvitenskap også!
# Mønsterdannelse under celle-differensiering

Harald Martens, Siren R Veflingstad, Erik Plahte, Magni Martens, Dominique Bertrand and Stig W Omholt (2009) The genotype-phenotype relationship in multicellular pattern-generating models - the neglected role of pattern descriptors.
BMC Systems Biology, 3:87 doi:10.1186/1752-0509-3-87

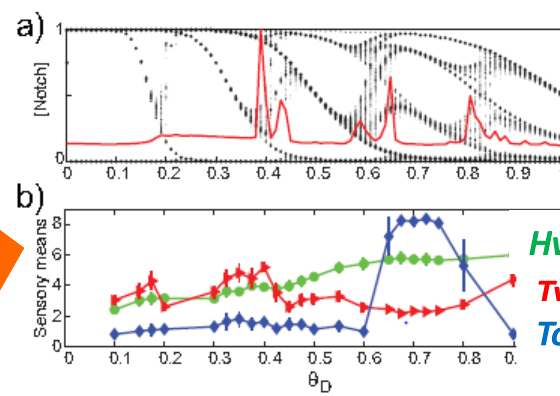



Matematisk modell av mønsterdannelse ved celle-differensiering

Datamaskin-simuleringer

Sensorisk deskriptiv profilering på Matforsk

Fant nye egenskaper i den matematiske modellen




Fordeling av **tetthet** og **beregnings-tid**

***Hvithet***

***Tykkelse***

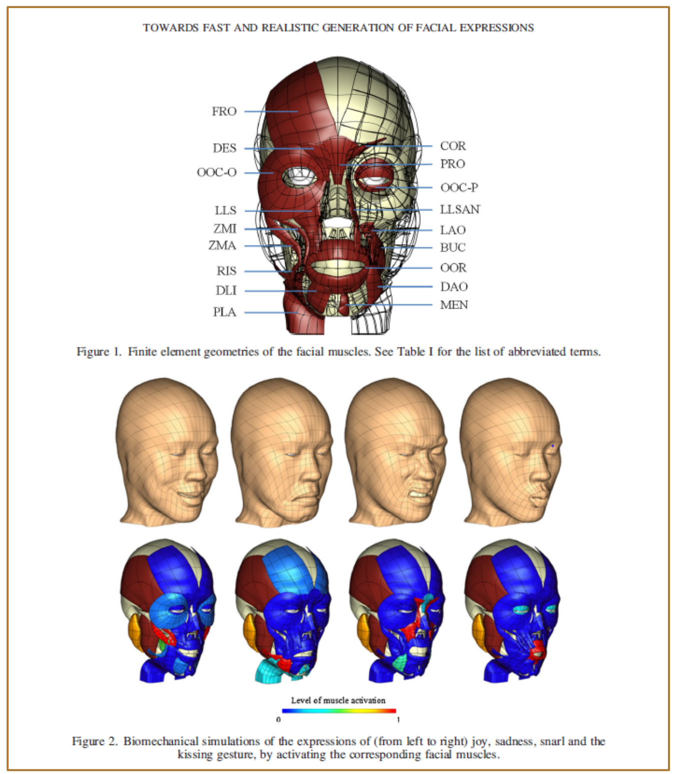
***Tohodethet***

# *Kroppen* har rytmer og harmonier

## 1: Verden har mønstre – finn dem

**Eksempel: Speed-up av matematisk formulering av ansikts-stemninger**

CPU beregningtid per ansiktsuttrykk:
**> 2 timer ⇒ < 0.01 sekund**




INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN BIOMEDICAL ENGINEERING
*Int. J. Numer. Meth. Biomed. Engng.* (2014)
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/cnm.2646

Emulating facial biomechanics using multivariate partial least squares surrogate models

Tim Wu[1,*,†], Harald Martens[2], Peter Hunter[1] and Kumar Mithraratne[1]

# Application: Speeding up a model

**Emulating facial biomechanics using multivariate partial least squares surrogate models**

Tim Wu[1,*], Harald Martens[2], Peter Hunter[1] and Kumar Mithraratne[1]

Article first published online: 6 MAY 2014

DOI: 10.1002/cnm.2646



Issue

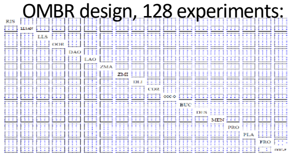
International Journal for Numerical Methods in Biomedical Engineering

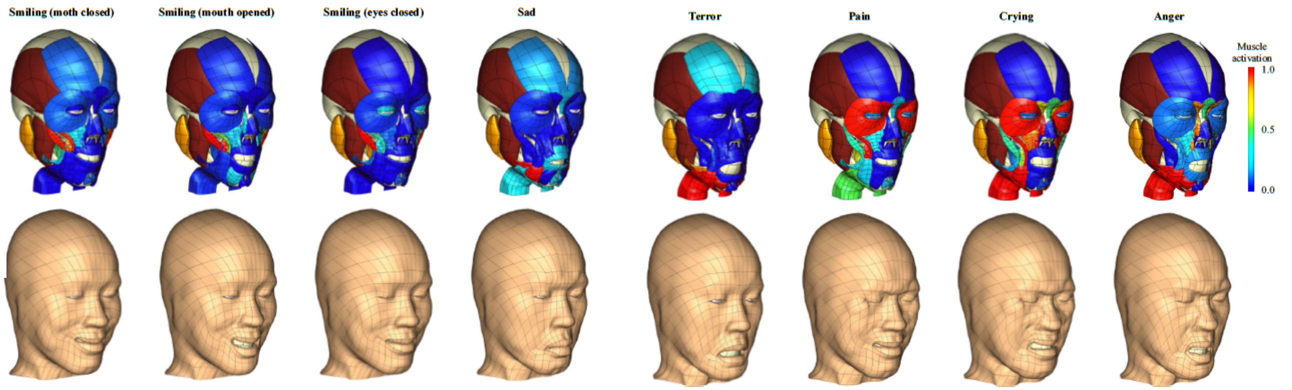
Early View (Online Version of Record published before inclusion in an issue)

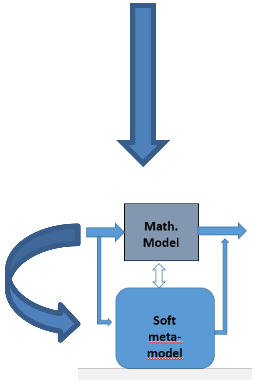
## Computational compaction of a large biological model via multivariate metamodellering

FEM of facial muscle groups
18 parameters, 4 levels of each.

OMBR design, 128 experiments:






Computational times:

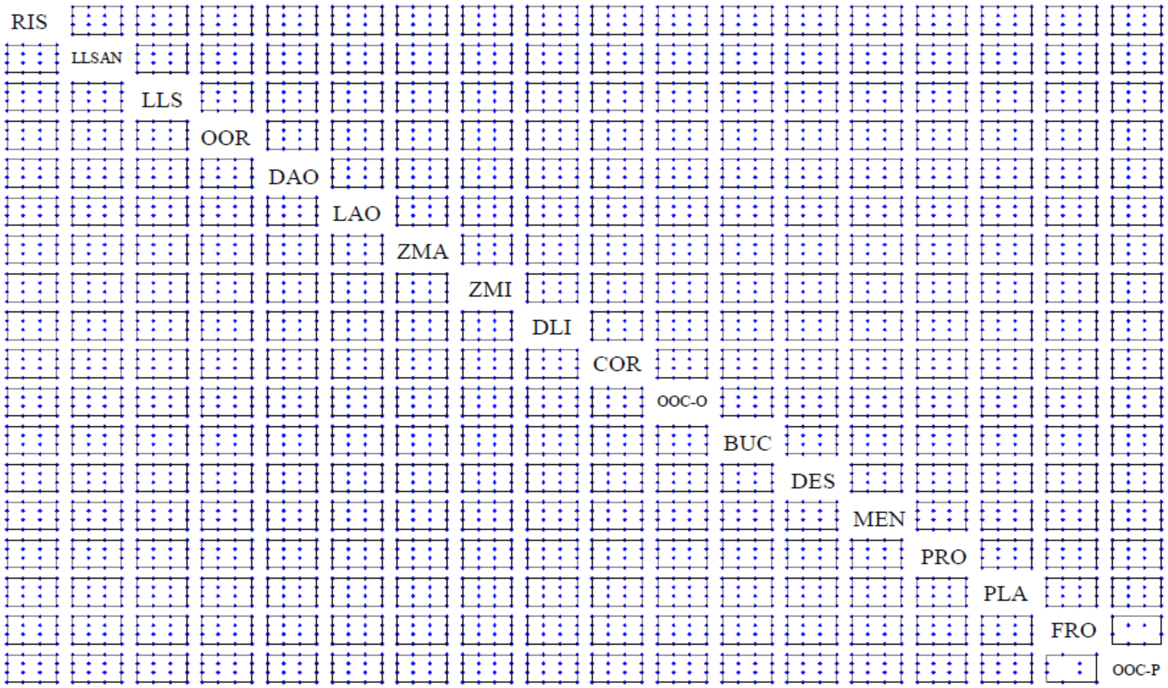
*Conventional FEM computation:* *> 120 minutes (average)*

*Multivariate metamodel (nonlin. PLSR):* *< 0.01 minute (average)*

Modern computer simulation design: 18 parameters, 4 levels of each

$4^{18} \approx 100\,000\,000\,000$ possible combinations

OMBR design, 128 experiments:



# Generic: Optimized Multi-level Binary Replacement design (OMBR design)

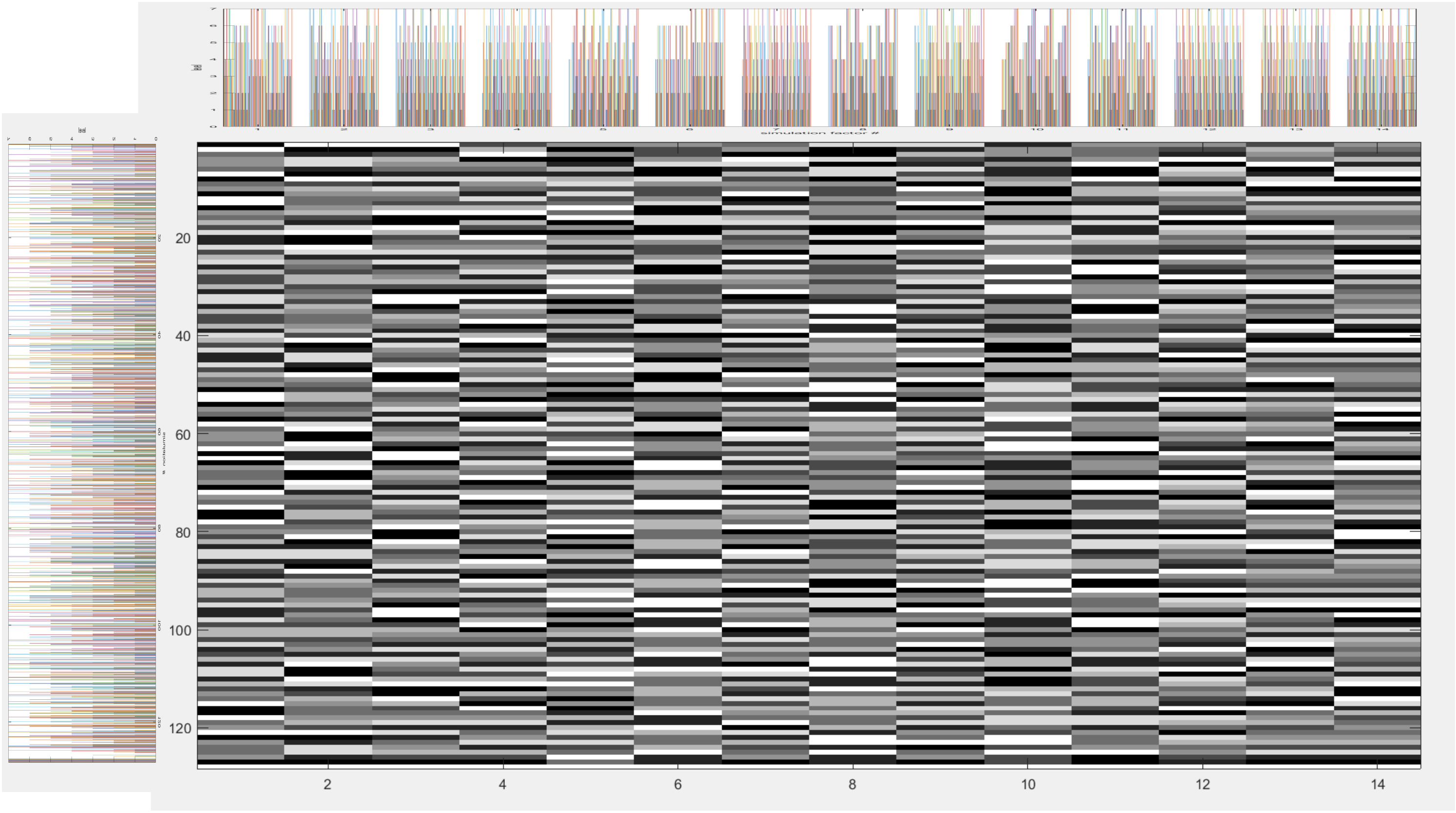
Give:

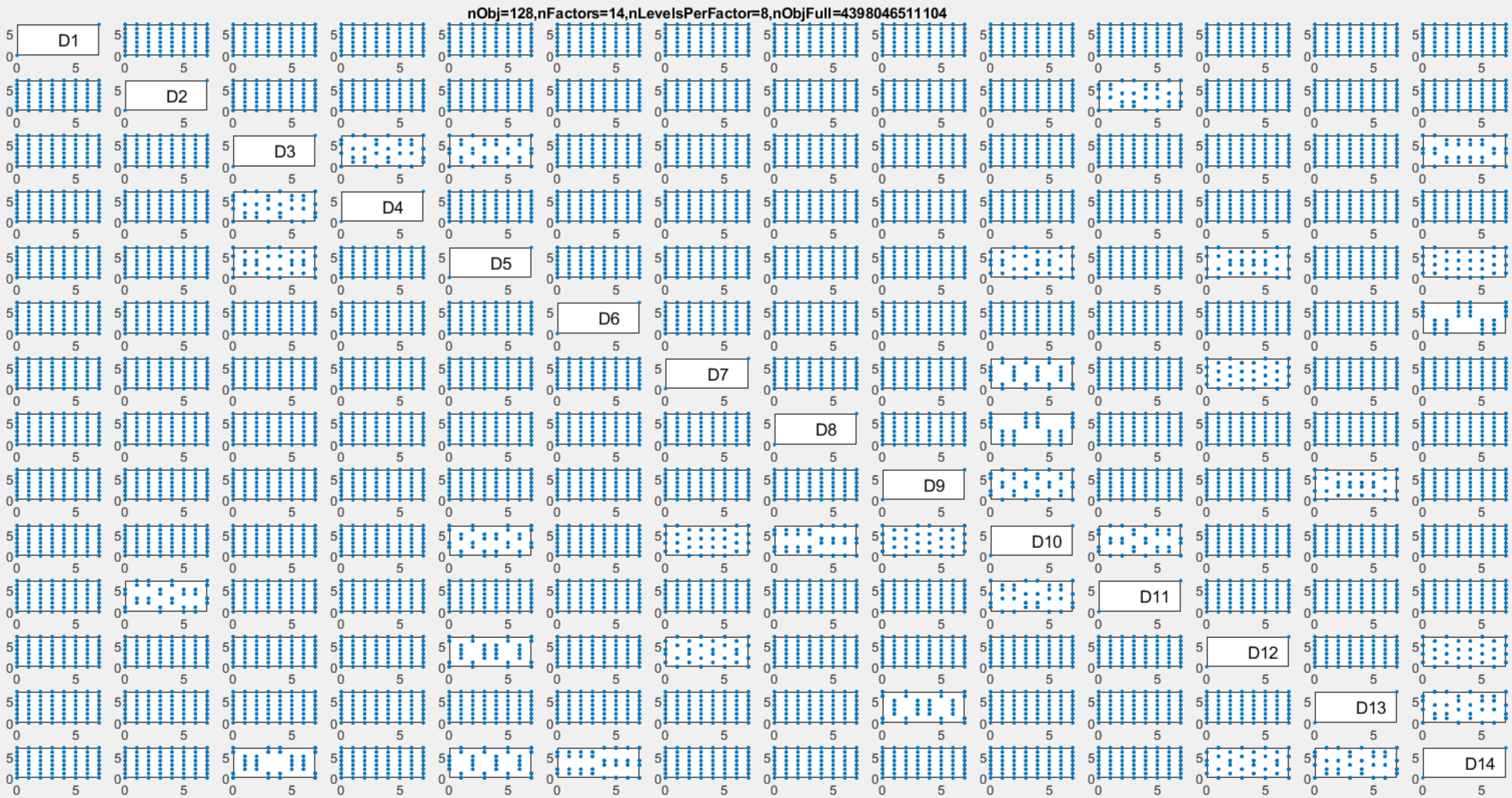
# of design factors ( input parameters to change, e.g. 4 or 18)

# of levels per design factor (e.g. 2, 4, 8 or 16)

# of simulations affordable (e.g.  64, 128,…,33000)

| Name | Date modified | Type | Size |
|---|---|---|---|
| OMBRDesign_12Factors_8Levels_256Simulations.mat | 28.02.2012 07:53 | Microsoft Access T... | 28 KB |
| OMBRDesign_12Factors_8Levels_512Simulations.mat | 28.02.2012 07:54 | Microsoft Access T... | 49 KB |
| OMBRDesign_12Factors_8Levels_1024Simulations.mat | 28.02.2012 07:55 | Microsoft Access T... | 91 KB |
| OMBRDesign_12Factors_8Levels_2048Simulations.mat | 28.02.2012 07:57 | Microsoft Access T... | 175 KB |
| OMBRDesign_12Factors_8Levels_4096Simulations.mat | 28.02.2012 07:59 | Microsoft Access T... | 343 KB |
| OMBRDesign_12Factors_8Levels_8192Simulations.mat | 28.02.2012 08:03 | Microsoft Access T... | 679 KB |
| OMBRDesign_12Factors_8Levels_16384Simulations.mat | 28.02.2012 08:10 | Microsoft Access T... | 1 351 KB |
| OMBRDesign_12Factors_8Levels_32768Simulations.mat | 28.02.2012 08:23 | Microsoft Access T... | 2 695 KB |
| OMBRDesign_12Factors_16Levels_64Simulations.mat | 28.02.2012 08:24 | Microsoft Access T... | 13 KB |
| OMBRDesign_12Factors_16Levels_128Simulations.mat | 28.02.2012 08:25 | Microsoft Access T... | 20 KB |
| OMBRDesign_12Factors_16Levels_256Simulations.mat | 28.02.2012 08:27 | Microsoft Access T... | 34 KB |
| OMBRDesign_12Factors_16Levels_512Simulations.mat | 28.02.2012 08:28 | Microsoft Access T... | 61 KB |
| OMBRDesign_12Factors_16Levels_1024Simulations.mat | 28.02.2012 08:30 | Microsoft Access T... | 115 KB |
| OMBRDesign_12Factors_16Levels_2048Simulations.mat | 28.02.2012 08:32 | Microsoft Access T... | 223 KB |
| OMBRDesign_12Factors_16Levels_4096Simulations.mat | 28.02.2012 08:35 | Microsoft Access T... | 439 KB |
| OMBRDesign_12Factors_16Levels_8192Simulations.mat | 28.02.2012 08:39 | Microsoft Access T... | 871 KB |
| OMBRDesign_12Factors_16Levels_16384Simulations.mat | 28.02.2012 08:45 | Microsoft Access T... | 1 735 KB |
| OMBRDesign_12Factors_16Levels_32768Simulations.mat | 28.02.2012 09:00 | Microsoft Access T... | 3 463 KB |
| OMBR-Design-8-4-128-27-Aug-2019.mat | 27.08.2019 09:55 | Microsoft Access T... | 6 KB |
| OMBR-Design-14-8-128-27-Aug-2019.mat | 27.08.2019 10:02 | Microsoft Access T... | 14 KB |

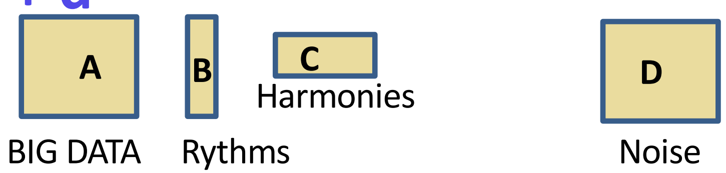

simulation factor #
20
40
60
80
100
120
2
4
6
8
10
12
14


nObj=128,nFactors=14,nLevelsPerFactor=8,nObjFull=4398046511104
D1
D2
D3
D4
D5
D6
D7
D8
D9
D10
D11
D12
D13
D14

# garbage

# Mighty, simple math:

- $21 = 3 \times 7$
- $21.1 = 3 \times 7 + 0.1$
- $a = b \times c + d$
- $31.1 = 3 \times 7 + 2 \times 5 + 0.1$
- $a = b_1 \times c_1 + b_2 \times c_2 + d$
- $\mathbf{a = b \times c + d}$




- $\mathbf{A = B \times C + D}$

(matrise-algebra, 1835)

$$21 = 3 \times 7$$
$$21.1 = 3 \times 7 + 0.1$$
$$a = b \times c + d$$
$$31.1 = 3 \times 7 + 2 \times 5 + 0.1$$
$$a = b_1 \times c_1 + b_2 \times c_2 + d$$
$$\mathbf{a} = \mathbf{b} \times \mathbf{c} + \mathbf{d}$$

$$\mathbf{A} = \mathbf{B} \times \mathbf{C} + \mathbf{D}$$

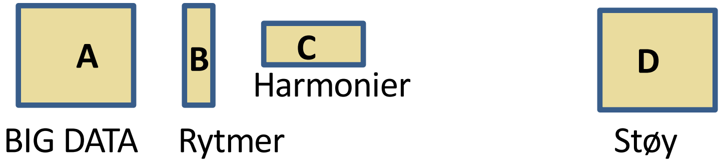
# Mighty, simple math:

- $21 = 3 \times 7$
- $21.1 = 3 \times 7 + 0.1$
- $a = b \times c + d$
- $\mathbf{A} = \mathbf{B} \times \mathbf{C} + \mathbf{D}$

- $\mathbf{A} = \mathbf{b_1} \times \mathbf{c_1} + \mathbf{b_2} \times \mathbf{c_2} + \ldots + \mathbf{b_m} \times \mathbf{c_m} + \ldots + \mathbf{b_M} \times \mathbf{c_M} + \mathbf{D_M}$

from data-table **A**, find structure(s) **B** × **C** and table **D**

Principal component analyse (PCA):
Mother of all multivariate metods!


A
B
C
Harmonier
D
BIG DATA
Rytmer
Støy

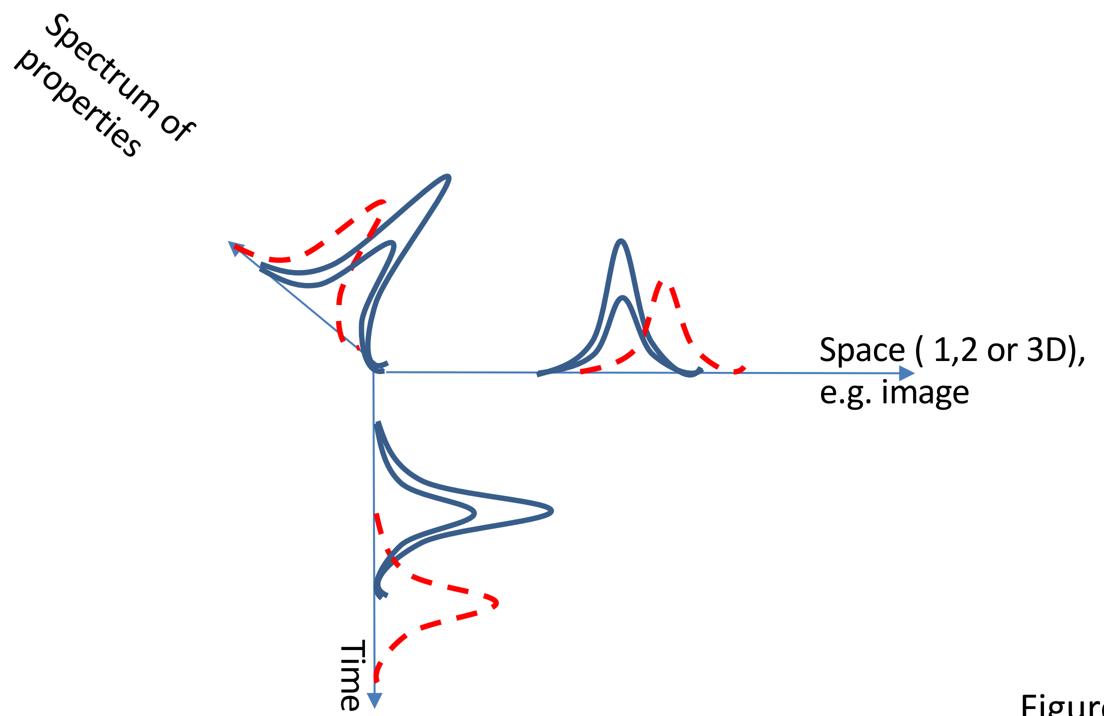
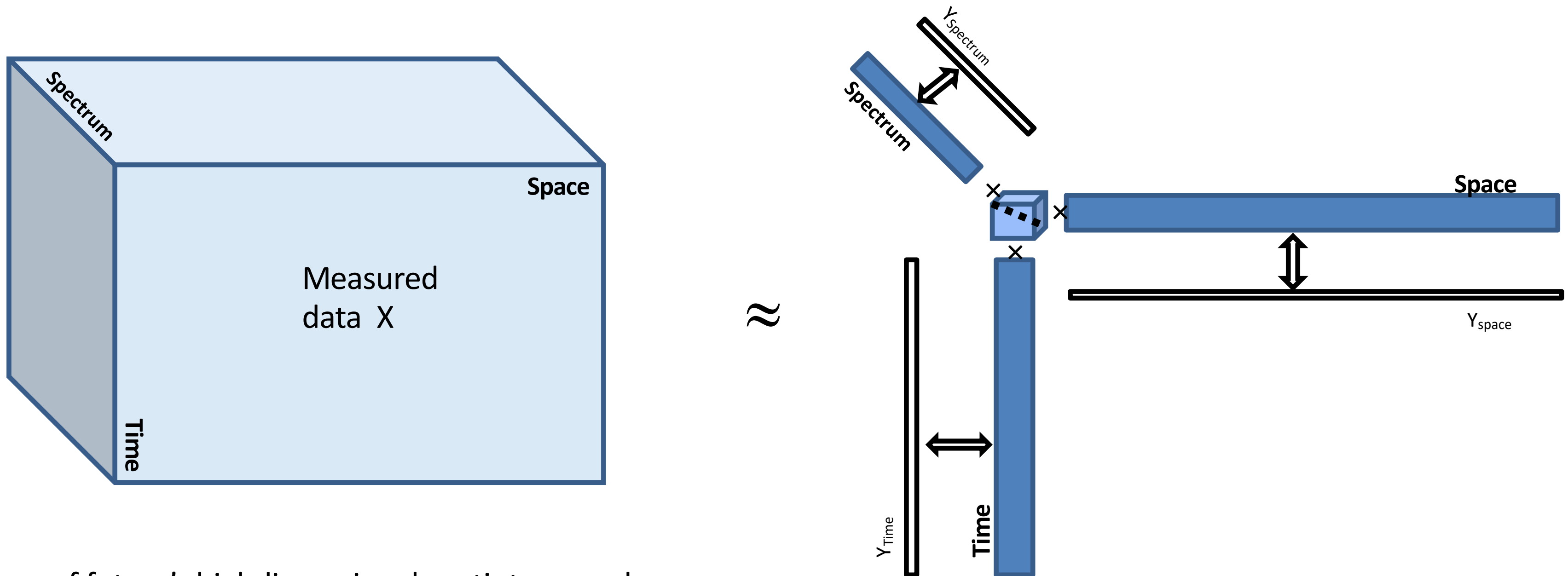


Figure 6a

Epistemology: measure position and intensity variation in time, space and properties, and extract interpretable essence by data modelling




Prototype of future's highdimensional spatiotemporal instrumentation:
**Hyperspectral NIR video**
1-2 GB of data/hr

Figure 7

# Mighty, simple math:

- $21 = 3 \times 7$
- $21.1 = 3 \times 7 + 0.1$
- $a = b \times c + d$

# Mighty, simple math:

- $21 \quad = 3 \times 7$
- $21.1 = 3 \times 7 \qquad + 0.1$
- $a \quad = b \times c \qquad + d$
- $31.1 = 3 \times 7 + 2 \times 5 \quad + 0.1$
- $a \quad = b_1 \times c_1 + b_2 \times c_2 \quad + d = \quad [b_1\ b_2] \times [c_1\ c_2]' + d$

# Mighty, simple math:

- $21 = 3 \times 7$
- $21.1 = 3 \times 7 + 0.1$
- $a = b \times c + d$
- $31.1 = 3 \times 7 + 2 \times 5 + 0.1$
- $a = b_1 \times c_1 + b_2 \times c_2 + d = [b_1\ b_2] \times [c_1\ c_2]' + d$
- $a = \mathbf{b} \times \mathbf{c} + d$ (vector algebra, Caspar Wessel 1797)

# Mighty, simple math:

- $21 \quad = 3 \times 7$
- $21.1 = 3 \times 7 \qquad + 0.1$
- $a \quad = b \times c \qquad + d$
- $31.1 = 3 \times 7 + 2 \times 5 \qquad + 0.1$
- $a \quad = b_1 \times c_1 + b_2 \times c_2 \quad + d = \quad [b_1\ b_2] \times [c_1\ c_2]' + d$
- $a \quad = \mathbf{b} \times \mathbf{c} \qquad + d$ (vector algebra, Caspar Wessel 1797)

- $\mathbf{A} \quad = \mathbf{b_1 \times c_1 \quad + \quad b_2 \times c_2 \quad + \ldots + \quad b_m \times c_m \quad + \ldots + \quad b_M \times c_M \quad + D_M}$

BIG DATA

- $\boxed{\mathbf{A}}$

# Mighty, simple math:

- $21 = 3 \times 7$
- $21.1 = 3 \times 7 + 0.1$
- $a = b \times c + d$
- $31.1 = 3 \times 7 + 2 \times 5 + 0.1$
- $a = b_1 \times c_1 + b_2 \times c_2 + d = [b_1\ b_2] \times [c_1\ c_2]' + d$
- $a = \mathbf{b} \times \mathbf{c} + d$ (vector algebra, Caspar Wessel 1797)

- $\mathbf{A} = \mathbf{b_1} \times \mathbf{c_1} + \mathbf{b_2} \times \mathbf{c_2} + \ldots + \mathbf{b_m} \times \mathbf{c_m} + \ldots + \mathbf{b_M} \times \mathbf{c_M} + \mathbf{D_M}$

BIG DATA Rythms Harmonies Noise

- $\boxed{\mathbf{A}} = \boxed{\mathbf{B}} \times \boxed{\mathbf{C}} + \boxed{\mathbf{D}}$ (matrix algebra, 1835)

# Mighty, simple math:

- $21 \quad = 3 \times 7$
- $21.1 = 3 \times 7 \qquad + 0.1$
- $a \quad = b \times c \qquad + d$
- $31.1 = 3 \times 7 + 2 \times 5 \qquad + 0.1$
- $a \quad = b_1 \times c_1 + b_2 \times c_2 \quad + d = \quad [b_1\ b_2] \times [c_1\ c_2\ ]' + d$
- $a \quad = \mathbf{b} \times \mathbf{c} \qquad + d$ (vector algebra, Caspar Wessel 1797)

- $\mathbf{A} \quad = \mathbf{b_1} \times \mathbf{c_1} \ + \ \mathbf{b_2} \times \mathbf{c_2} \ + \ldots + \ \mathbf{b_m} \times \mathbf{c_m} \ + \ldots + \ \mathbf{b_M} \times \mathbf{c_M} \ + \mathbf{D_M}$

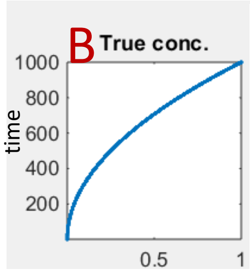
BIG DATA Rythms Harmonies Noise

- $\boxed{\mathbf{A}} = \boxed{\mathbf{B}} \times \boxed{\mathbf{C}} + \boxed{\mathbf{D}}$ (matrix algebra, 1835)

Principal component analyse (PCA), Mother of all multivariate metods:

From input data-table **A**, find structure(s) **B** × **C** and residual table **D**

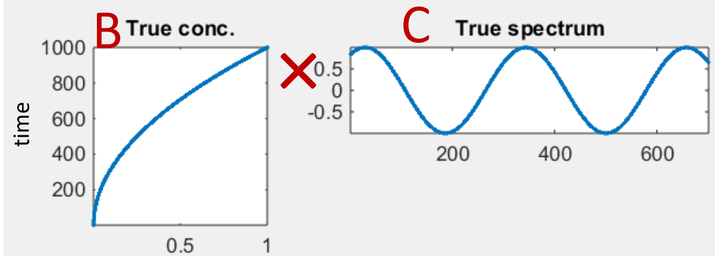
**Might, simple math**




A causal phenomenon's time-dependent development

## *How Quantitative Big Data are often generated*

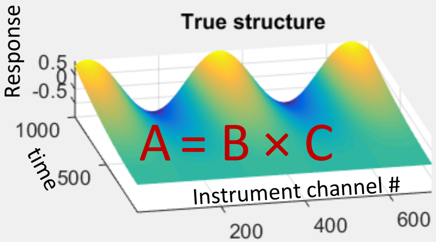
**Might, simple math**



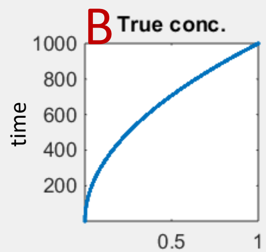

A causal phenomenon’s time-dependent development

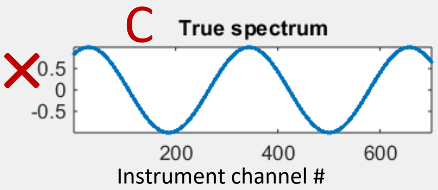
Its multi-channel property profile

## *How Quantitative Big Data are often generated*




True properties

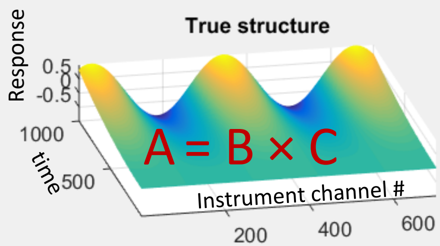


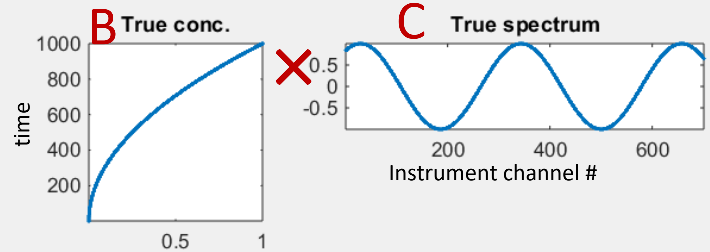




A causal phenomenon's time-dependent development

Its multi-channel property profile

***How Quantitative Big Data are often generated***



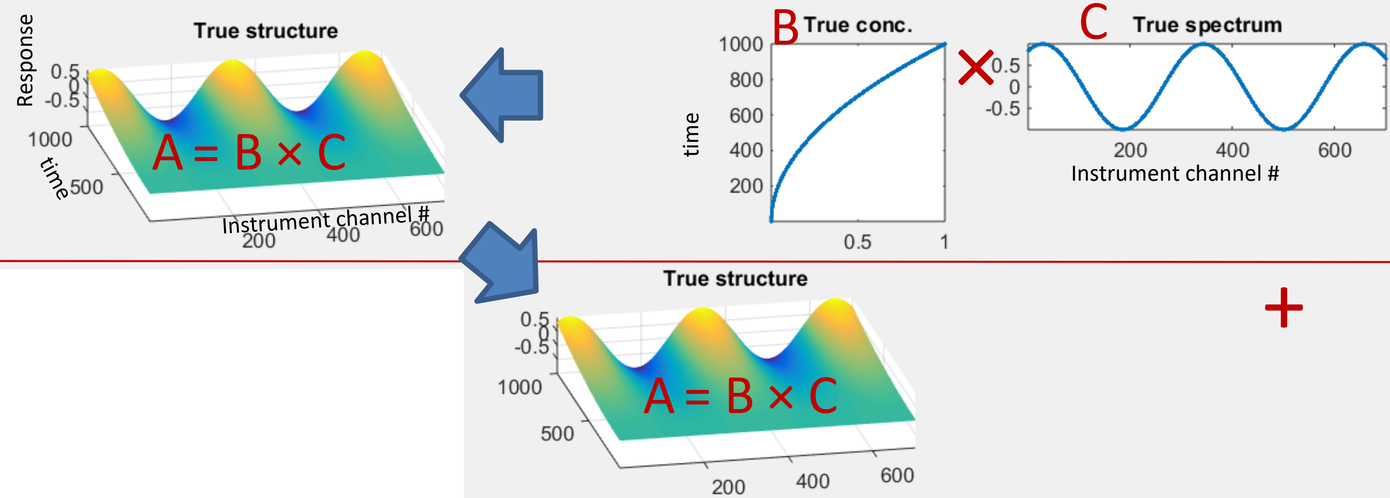

True properties

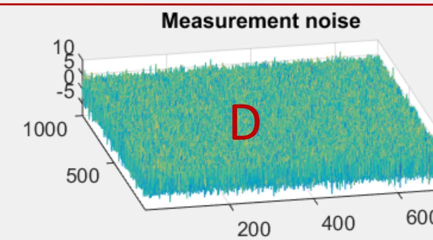
A causal phenomenon's time-dependent development

Its multi-channel property profile

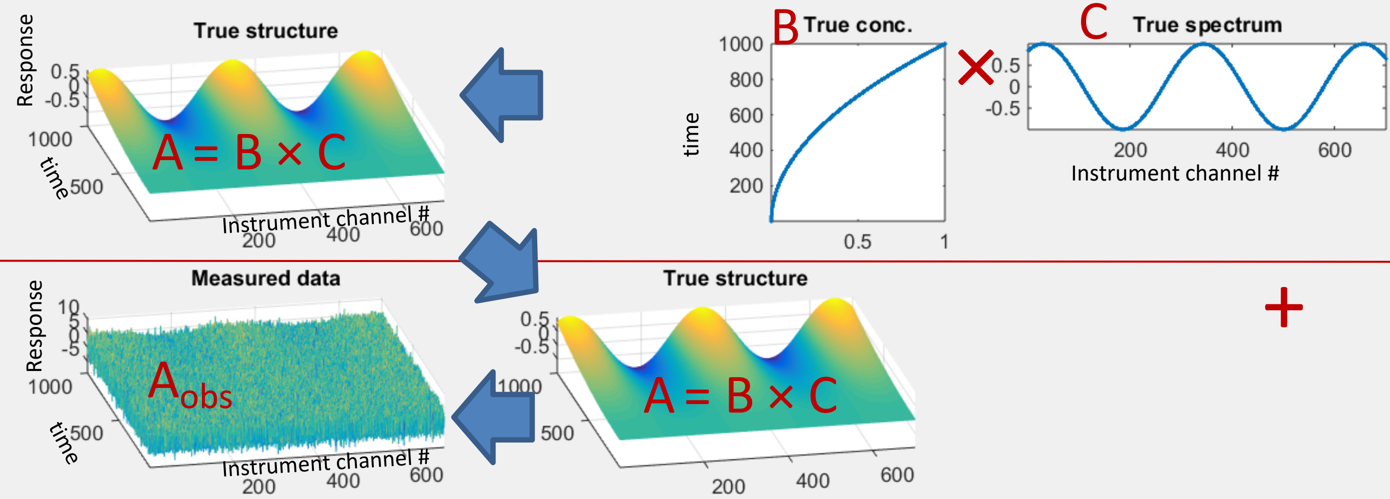
***Vector algebra***
***First: Caspar Wessel from Vestby, 1797***

## *How Quantitative Big Data are often generated*



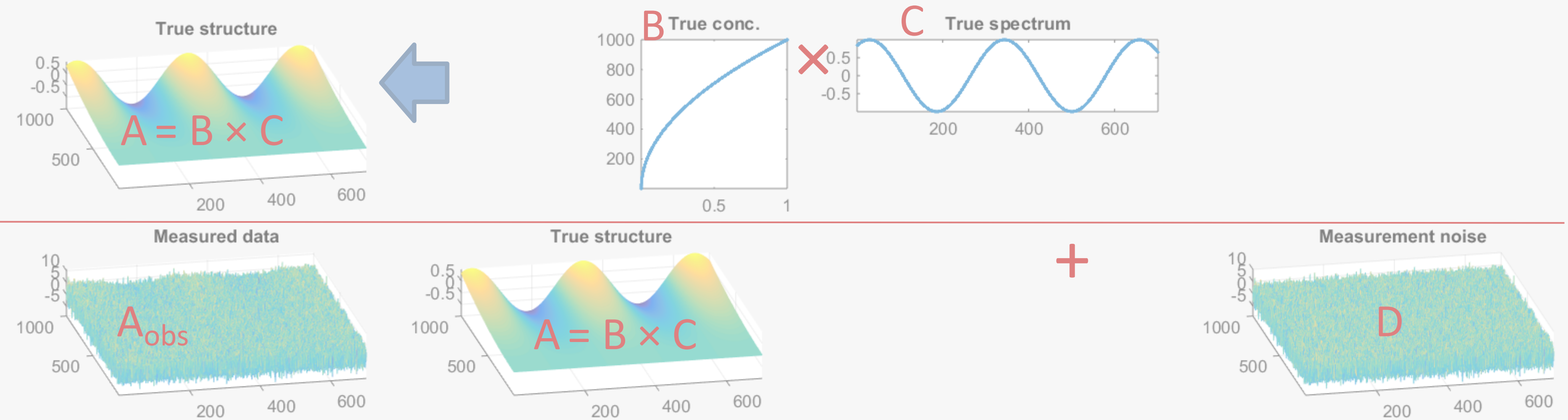

# *How Quantitative Big Data are often generated*



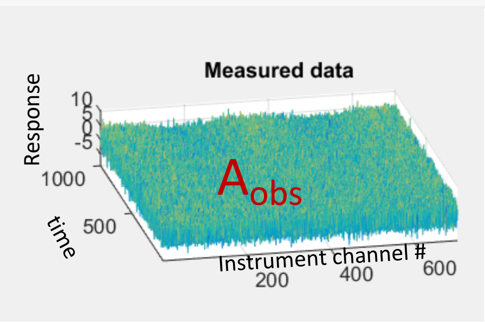

CAUSAL MATH MODEL:

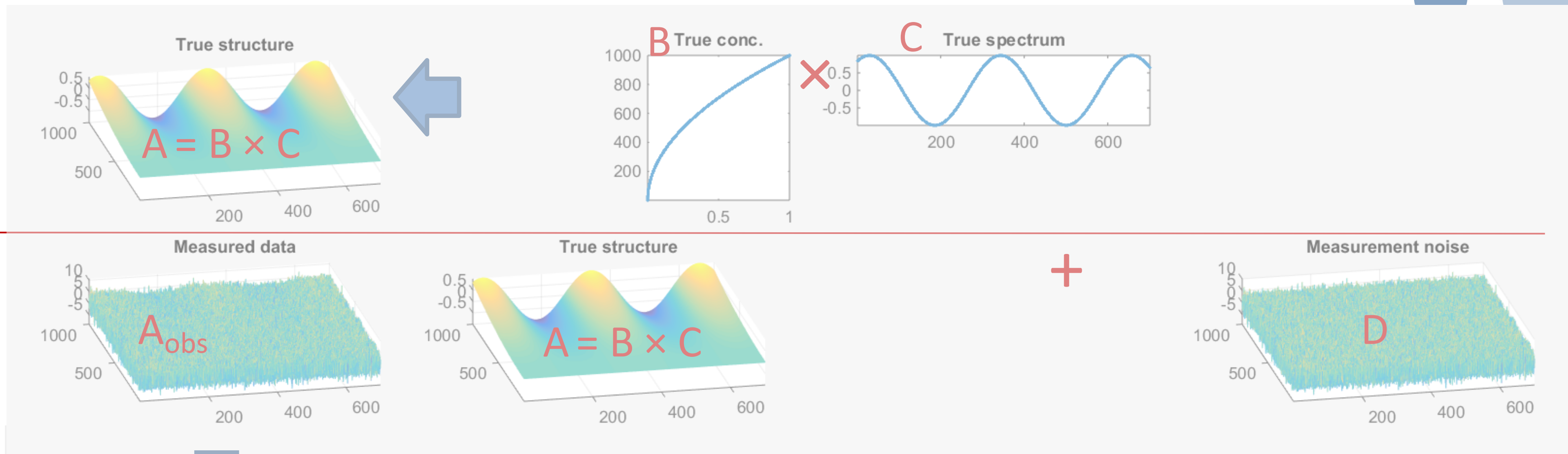
$A_{obs} = B \times C + D$

***How Quantitative Big Data are often generated***

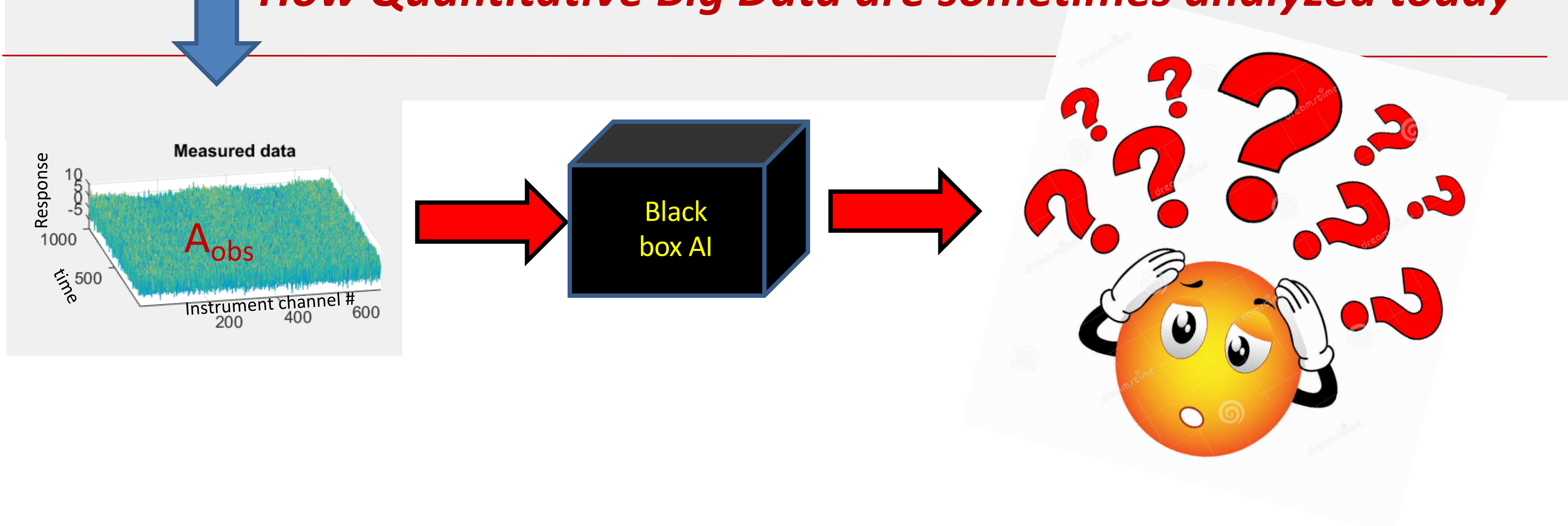


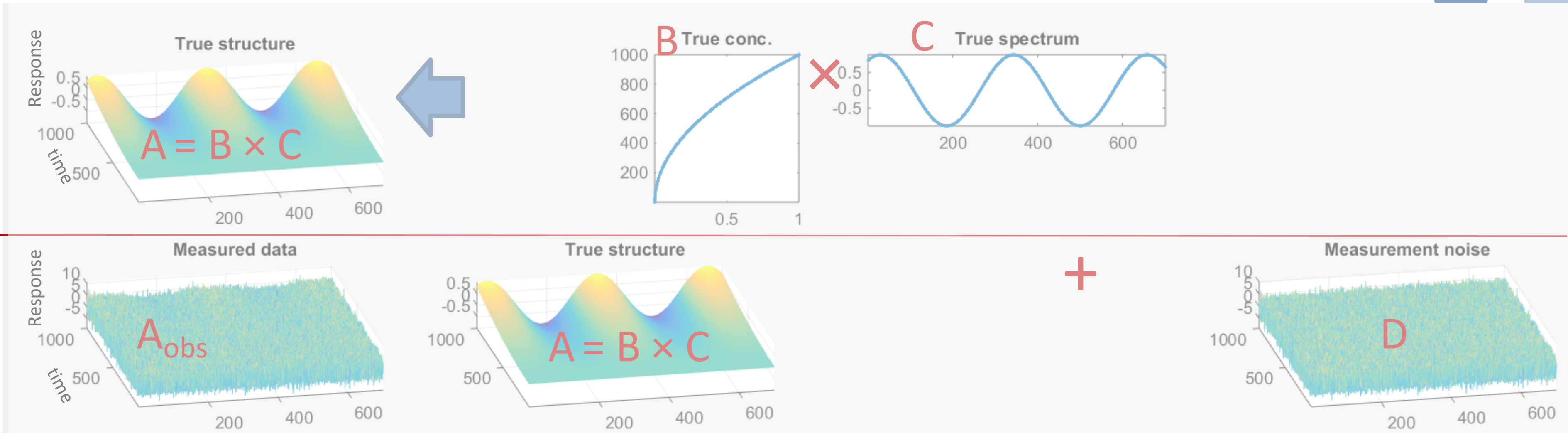

## *How Quantitative Big Data are NOT analyzed*



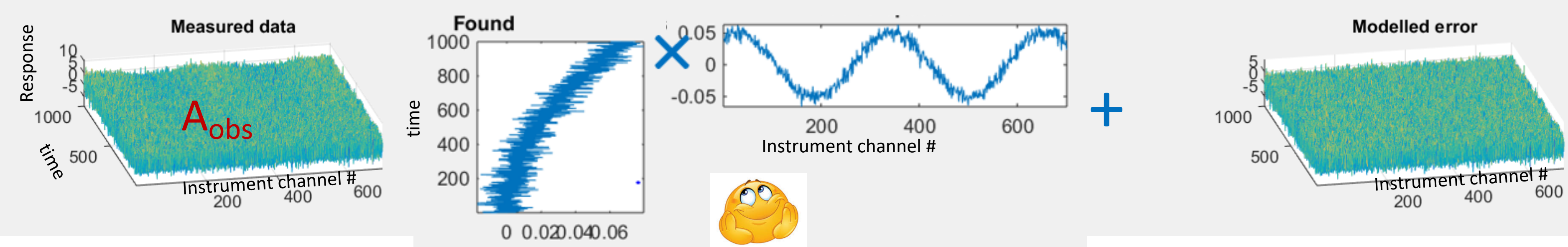

## *How Quantitative Big Data are sometimes analyzed today*

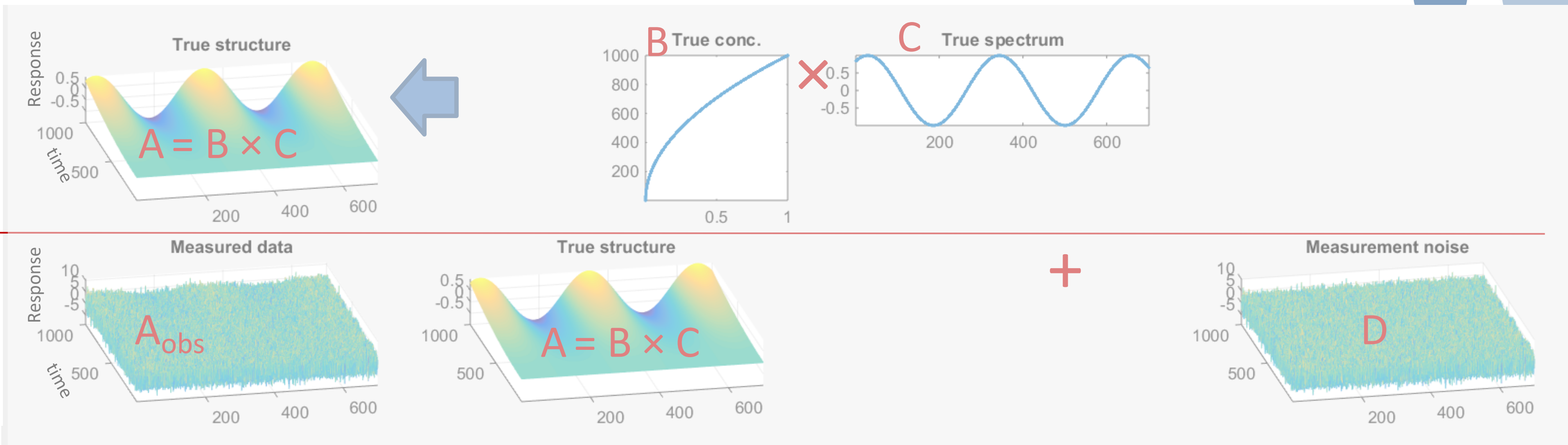





## *How Quantitative Big Data may be analyzed*

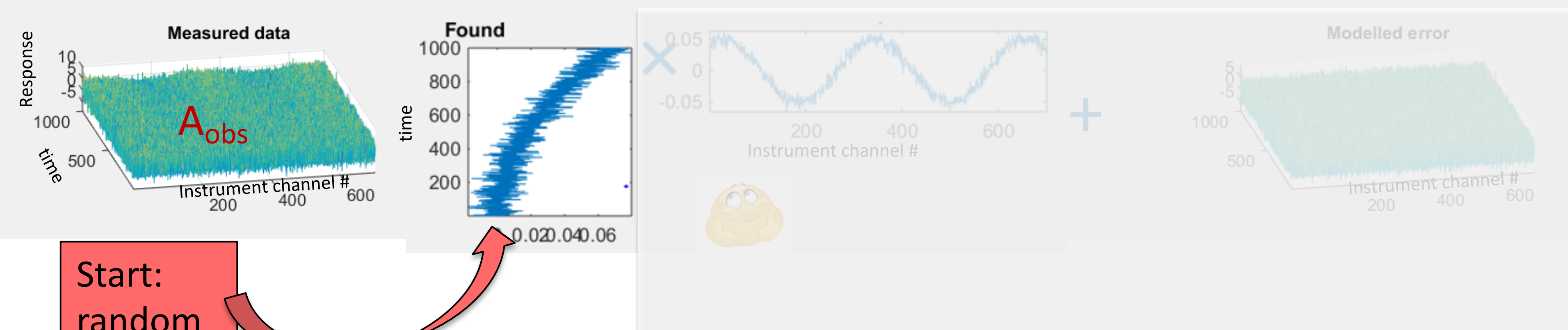
### Multivariate analysis:



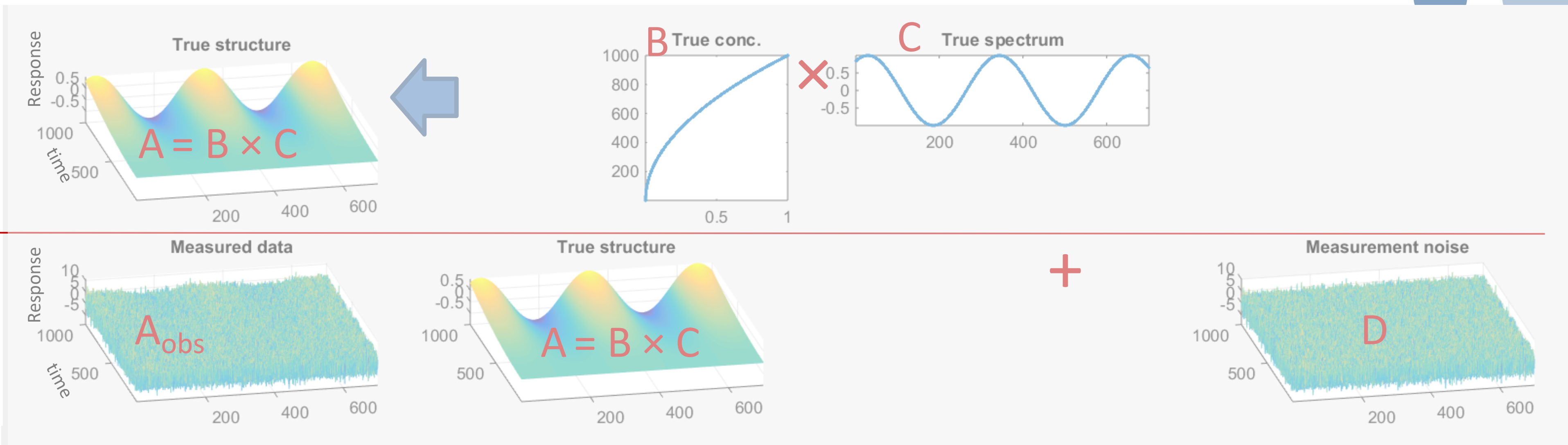



## *One way to do it: the NIPALS algorithm*

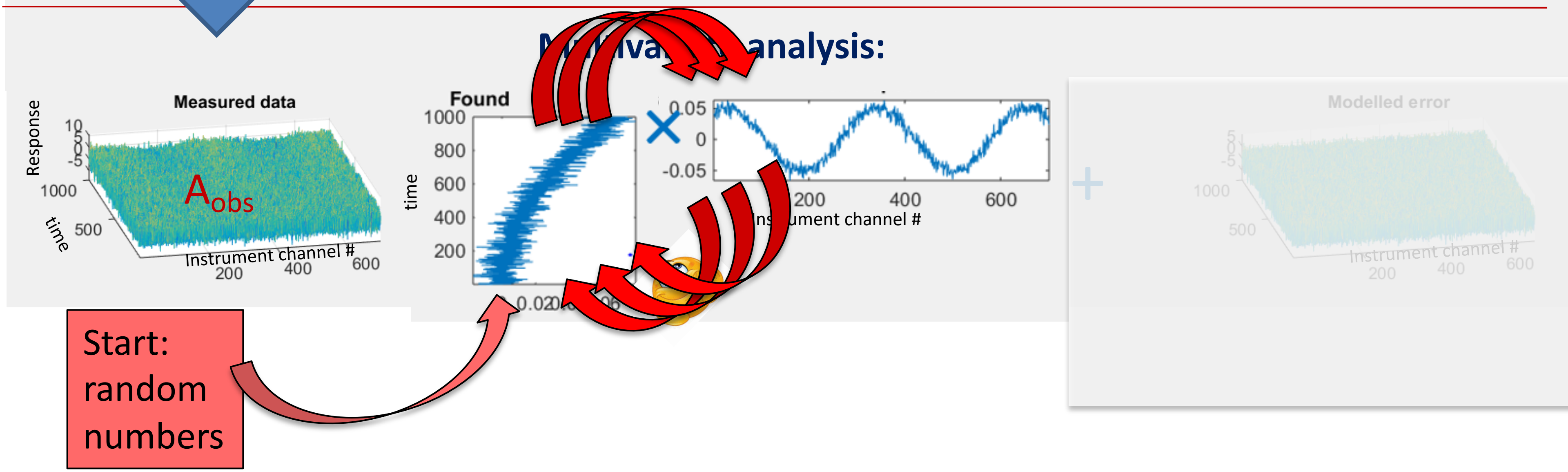
## Multivariate analysis:

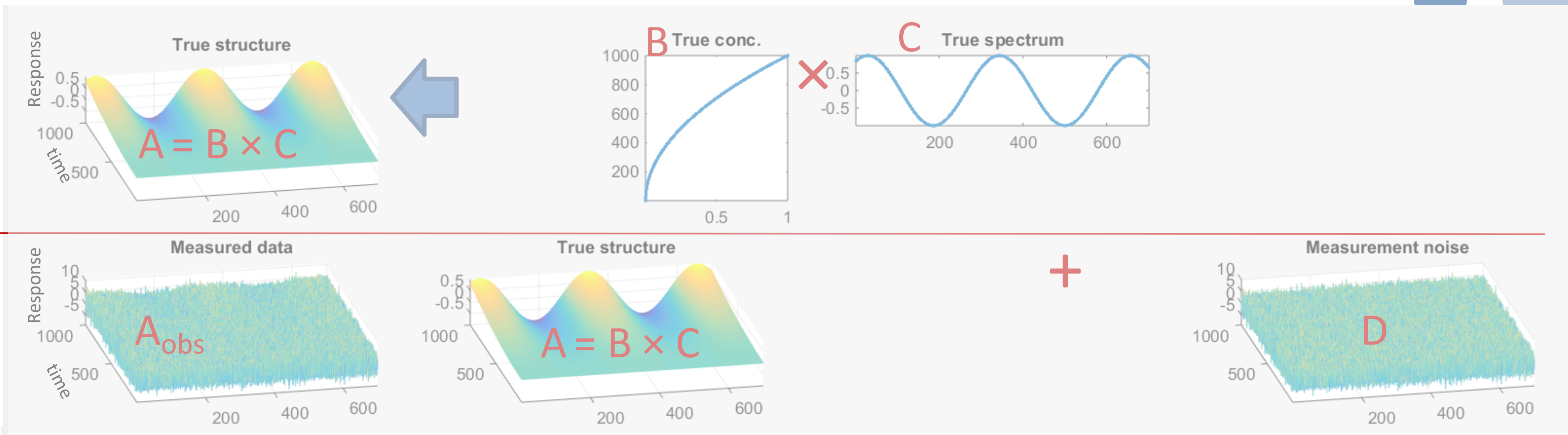



Start:
random
numbers

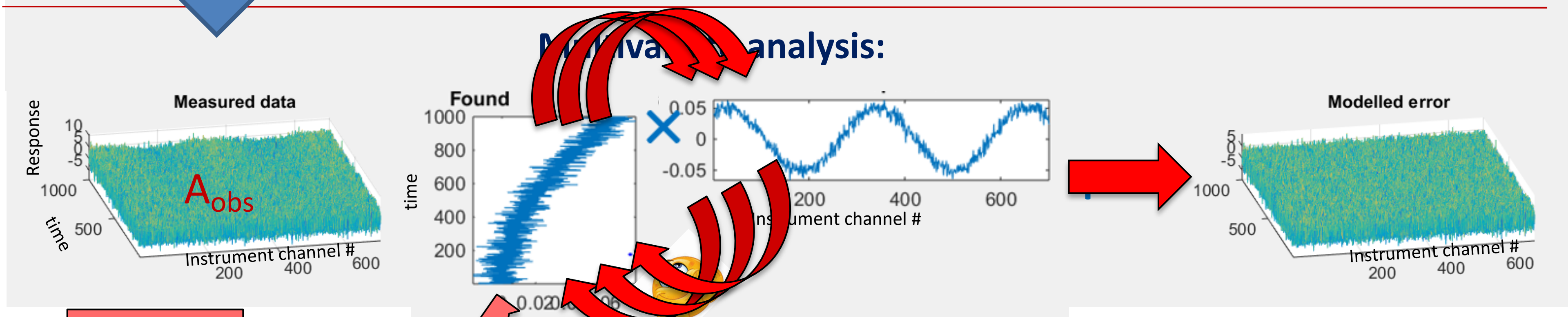



## *One way to do it: the NIPALS algorithm*






## *One way to do it: the NIPALS algorithm*




Start:
random numbers

# ANalysis Of VAriance (ANOVA)

- ANOVA separates data into contributions from structure and noise
- Data = Structure + Noise
  $SS_{Tot} = SS_{Reg} + SS_{Err}$
  Total variation = Modelled + Not modelled
- Success depends on quality of the data

# ANOVA output (1/2)

- Summary-section:
  - Model ($SS_{Reg}$):
    - Contribution from all terms in model
    - DF given by number of terms (parameters)
  - Error ($SS_{Err}$):
    - Non-modelled variation or noise
    - DF given by number of runs – number of terms - 1

  - Significance of model estimated from

    $F\text{-}ratio = MS_{Reg} / MS_{Err}$

# ANOVA output (2/2)

- Variables-section
  - The significance of each model parameter is estimated
- Model check-section
  - Sums the contribution from linear terms, interaction terms, etc.
- Lack of Fit-section
  - Total error may be divided into
    - Pure error: Spread between replicates
    - Lack of fit: Modelled values vs. Mean of replicates

# ANOVA Table

**Analysis of variance table [Partial sum of squares - Type III]**

| **Source** | **Sum of Squares** | **df** | **Mean Square** | **F Value** | **p-value Prob > F** | |
|---|---|---|---|---|---|---|
| Model | 5535.81 | 5 | 1107.16 | 56.74 | < 0.0001 | significant |
| *A-Temperature* | *1870.56* | *1* | *1870.56* | *95.86* | *< 0.0001* | |
| *B-Concentration* | *390.06* | *1* | *390.06* | *19.99* | *0.0012* | |
| *C-Stir Rate* | *855.56* | *1* | *855.56* | *43.85* | *< 0.0001* | |
| *AB* | *1314.06* | *1* | *1314.06* | *67.34* | *< 0.0001* | |
| *AC* | *1105.56* | *1* | *1105.56* | *56.66* | *< 0.0001* | |
| Residual | 195.12 | 10 | 19.51 | | | |
| Total | 5730.94 | 15 | | | | |

# Constructing a 2-level Fractional Factorial design: Confounding

- Example: Constructing the $2^{4-1}$ Design from a $2^3$ Design
- Write out the full $2^3$ Design

| A | B | C | AB | AC | BC | ABC |
|---|---|---|---|---|---|---|
| - | - | - | + | + | + | - |
| + | - | - | - | - | + | + |
| - | + | - | - | + | - | + |
| + | + | - | + | - | - | - |
| - | - | + | + | - | - | + |
| + | - | + | - | + | - | - |
| - | + | + | - | - | + | - |
| + | + | + | + | + | + | + |

Define I = ABCD

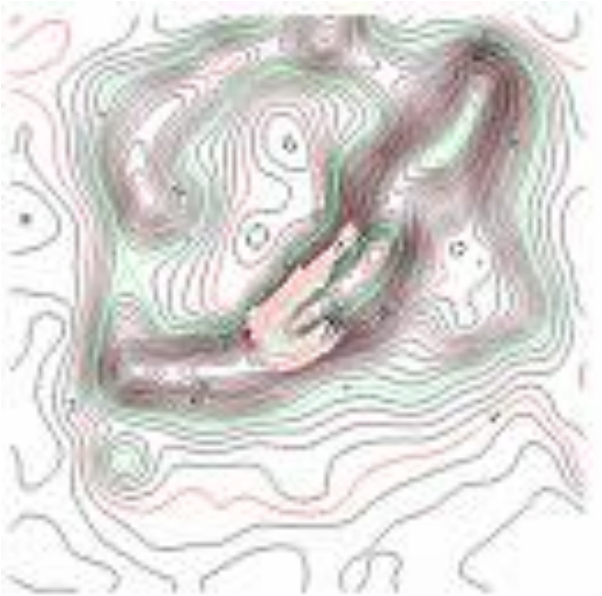
and

let D = ABC

Now reconstruct the design table with the above information

# Optimization Designs

- Objective
  - Model the variations of the responses with accuracy, so as to know the precise shape of the response surface, and (optionally) find optimum values
- Problem formulation
  - Include main effects
  - Include interactions
  - Include squared (and/or cubic) terms
- Designs
  - Central Composite designs
  - Box-Behnken design

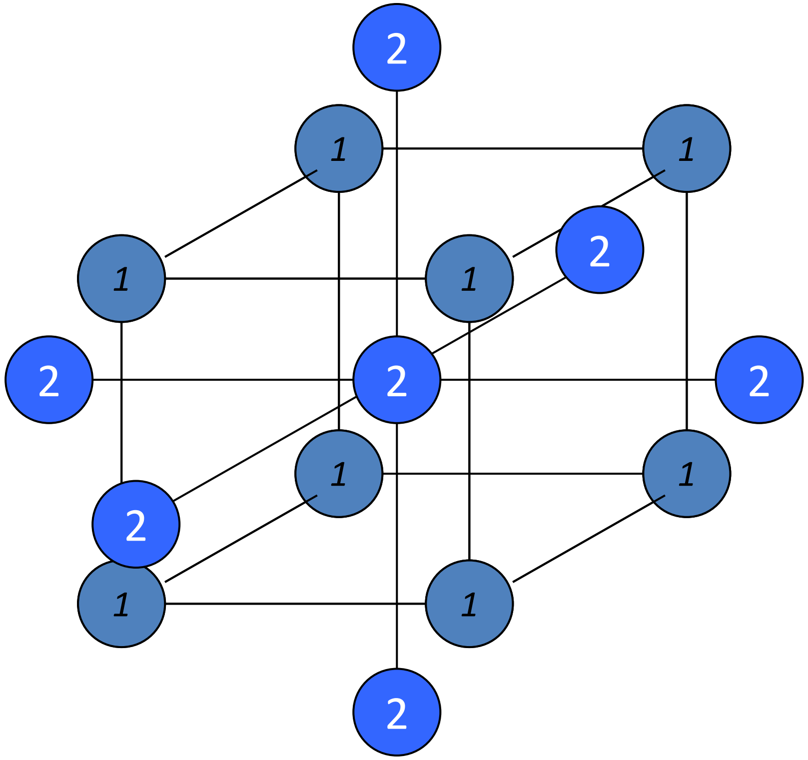


# Extending a full factorial

- A full factorial 2-level design can be extended to a ***Central Composite Design*** by adding star points

1: Factorial Points

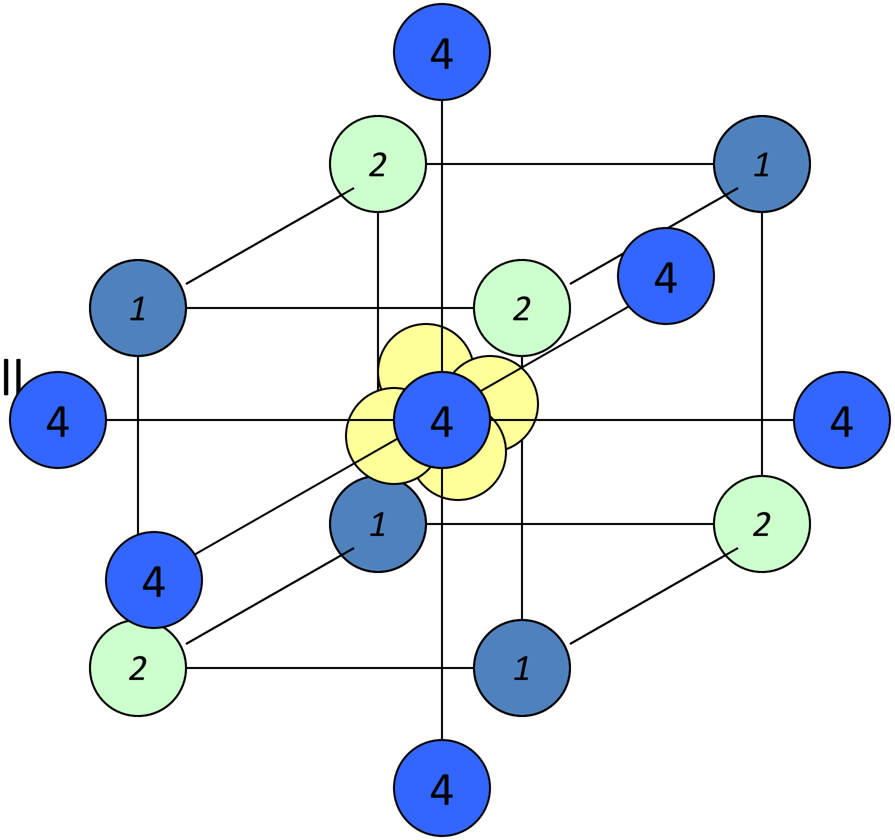
2: Star or Axial Points



# Central Composite Designs (CCD)

Composite of a factorial (linear) design and an optimization (quadratic) design

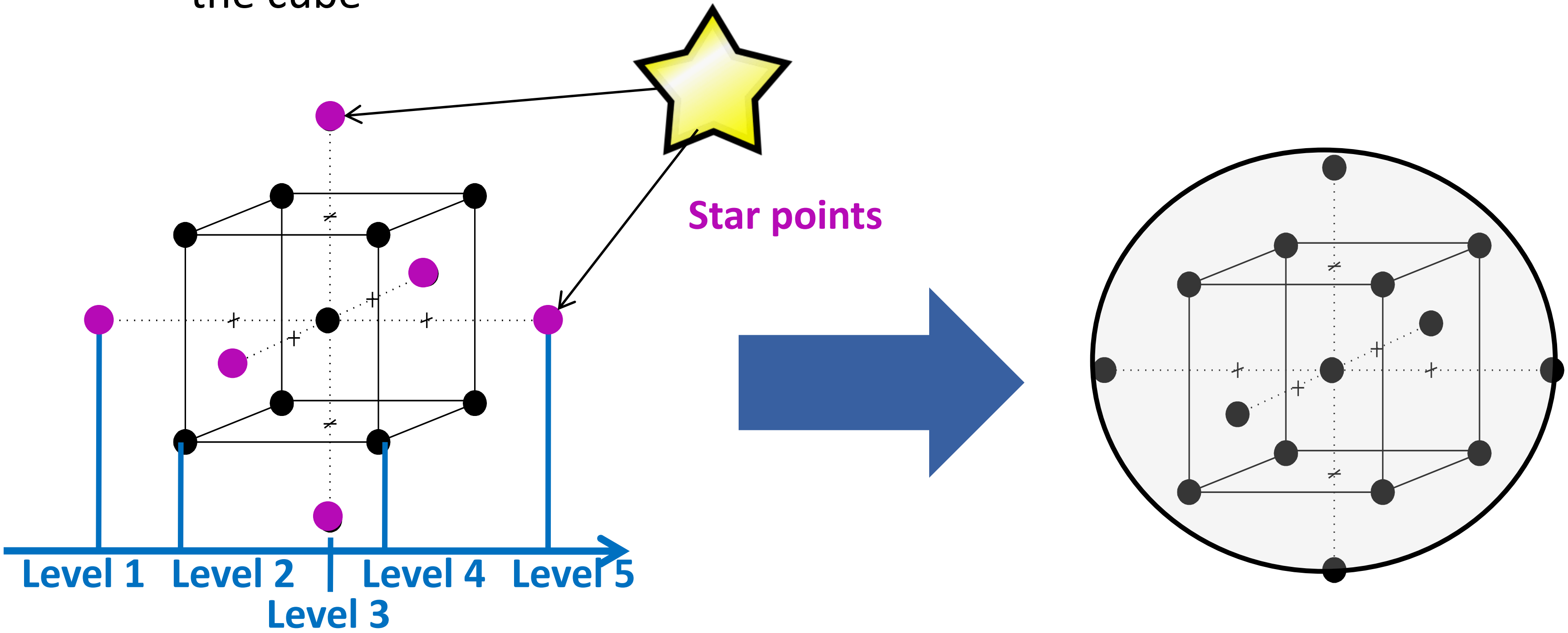
- Good for modelling a response surface
- 5 levels for each variable
- Can be built as an extension of a full factorial
- Additional points (blue) are called axial (or star) points



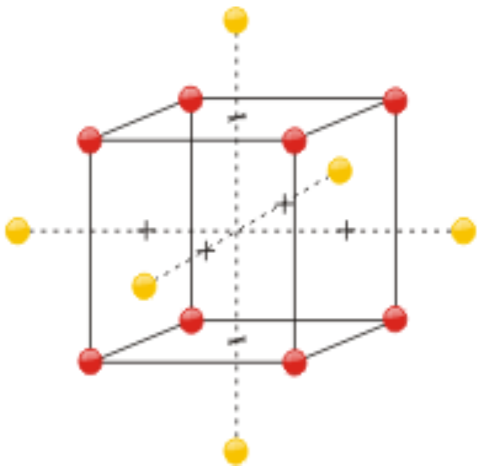

1. Fractional factorial, 2. Full factorial, 3. Centre points, 4. Axial points

# Central Composite Designs

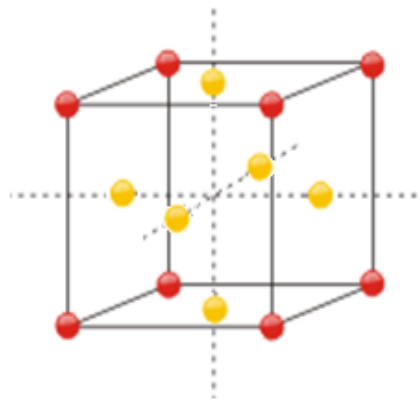
- Central Composite Circumscribed (CCC)
  - All points are on a sphere. The sphere encompasses the corners of the cube



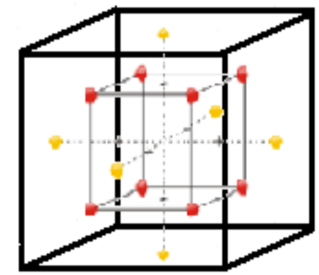
# Types of CCD



Circumscribed central composite design (CCC)



Faced central composite design (CCF)



Inscribed central composite design (CCI)

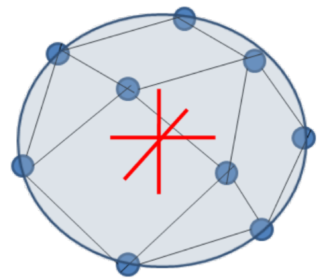
The black cube represents the user defined design space.

The red dots the cube points.

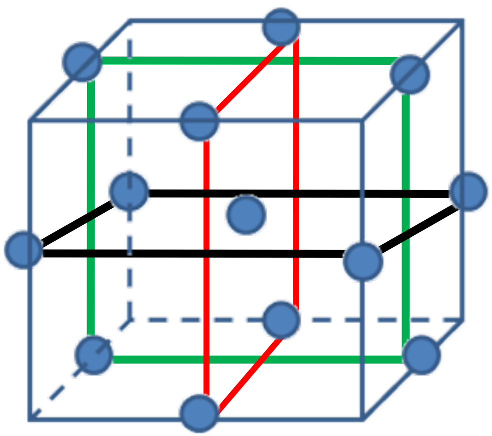
The yellow dots the star points.

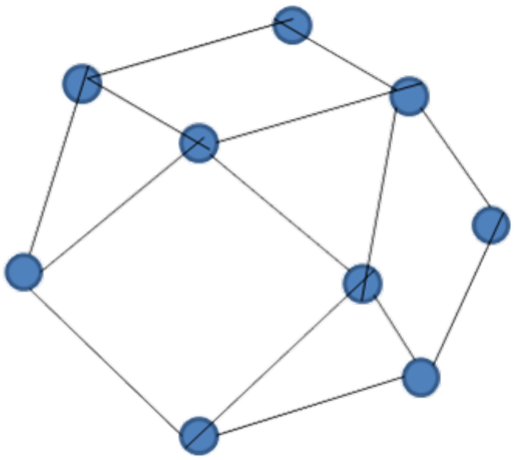
# Box-Behnken Designs

- 3 levels for each variable
- Slightly fewer experiments than CCD
- Extreme situations are avoided (cut-off corners)



Rotatable Design





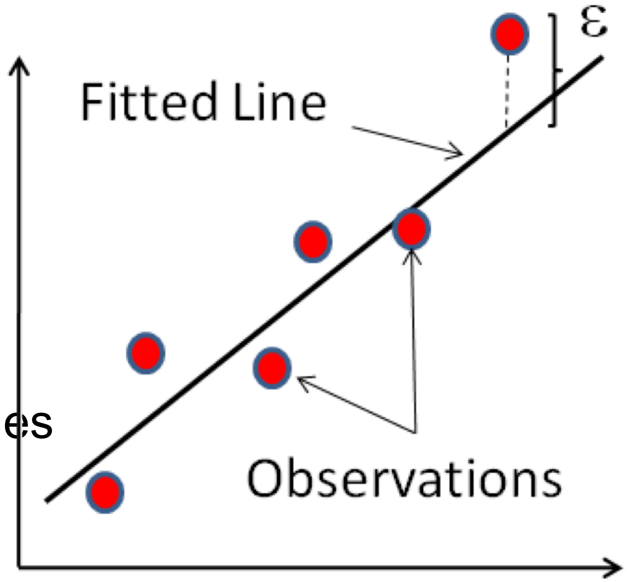
# Regression analysis

- Find a linear relationship between responses (y) and the design variables (x)

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n + \varepsilon$$

  – One X variable: Fit a line
  – Two X variables: Fit a plane
  – More than two: Fit a hyperplane

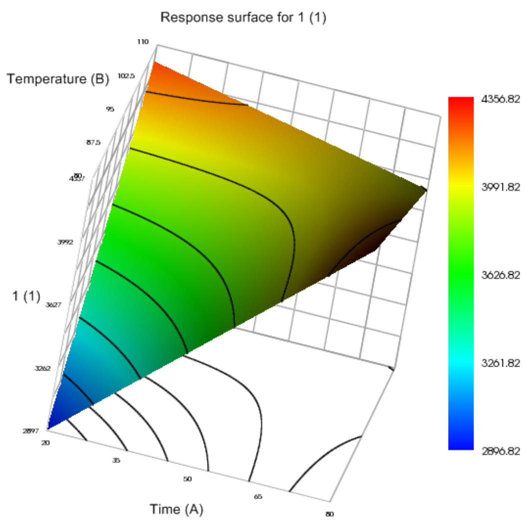
- The line (or plane) should lie as close as possible to the observations
  – Projections on the plane are called fitted values
  – Each observation has a residual (ε)



# Response Surface Methodology

- Purpose
  - Closely approximate the true shape of the response surface
- Quadratic model

$$y = b_0 + \sum b_i x_i + \sum b_i x_i^2 + \sum \sum b_{ij} x_i x_j + \varepsilon$$



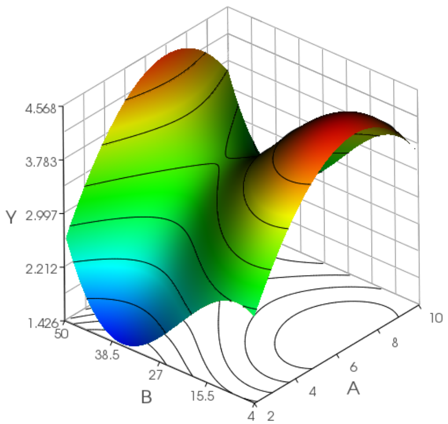

- Method: MLR
- Associated methods
  - ANOVA
  - Plots
- Use the model to
  - Predict the response value(s) for any combination of the the experimental region
  - Find the variable settings that give desired response value(s) in the experimental region (optimization)

# Response Surface ANOVA

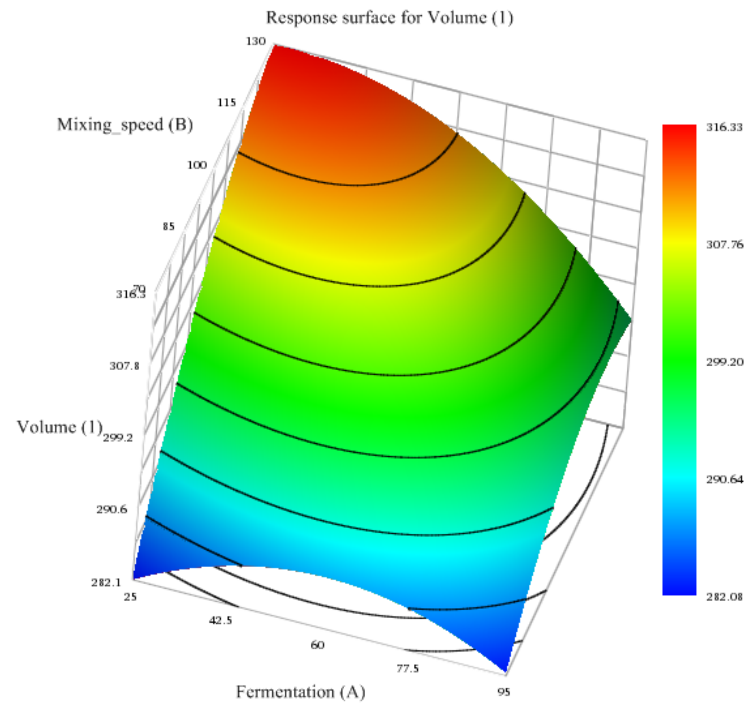
| ANOVA Table - Halt | | | | | |
|---|---|---|---|---|---|
| ANOVA | SS | DF | MS | F-ratio | p-value |
| **Summary** | | | | | |
| Model | 75.3750 | 3.0000 | 25.1250 | 240.5585 | 0.0000 |
| Error | 0.6267 | 6.0000 | 0.1044 | | |
| Corr. total | 76.0091 | 10.0000 | | | |
| **Variables** | | | | | |
| Temp (A) | 66.1250 | 1.0000 | 66.1250 | 633.1117 | 0.0000 |
| Feu (B) | 6.1250 | 1.0000 | 6.1250 | 58.6436 | 0.0003 |
| Bele (C) | 3.1250 | 1.0000 | 3.1250 | 29.9202 | 0.0016 |
| **Model check** | | | | | |
| Mean | | | | | |
| Linear | 75.3750 | 3.0000 | 25.1250 | 240.5585 | 0.0000 |
| Interaction 2 | 0.3750 | 3.0000 | 0.1250 | 1.1968 | 0.3879 |
| Interaction 3 | | | | | |
| Quadratic | 0.0074 | 1.0000 | 0.0074 | 0.0711 | 0.7987 |
| Cubic | 0.1250 | 1.0000 | 0.1250 | 1.1968 | 0.3159 |
| Total | 75.8824 | 10.0000 | 7.5882 | 72.6534 | 0.0000 |
| **Lack of fit** | | | | | |
| Lack of fit | 0.5000 | 4.0000 | 0.1250 | 1.9737 | 0.3634 |
| Pure Error | 0.1267 | 2.0000 | 0.0633 | | |
| Error | 0.6267 | 6.0000 | 0.1044 | | |
| **Quality** | | | | | |
| Method used | design | | | | |
| R-square | 0.9917 | | | | |
| Adjusted R-squa | 0.9881 | | | | |
| R-square predic | 0.9747 | | | | |
| S | 0.3232 | | | | |
| Mean | 6.3909 | | | | |
| C.V. in % | 5.0568 | | | | |
| PRESS | 1.9197 | | | | |

# Contour and surface plots

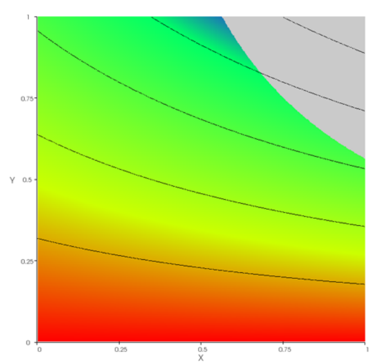
## Find max or min point




## Find stable region

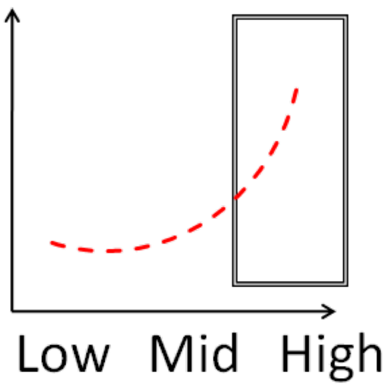


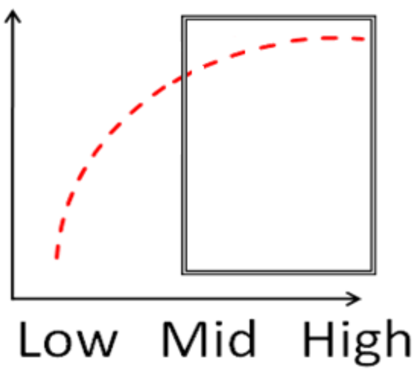

## Set constraints



# Ranges of variation for optimization

- Purposes:
  - Come as close as possible to the optimum
  - Describe response surface precisely
  - Ensure adequate variation in response
- Narrow range is often preferable



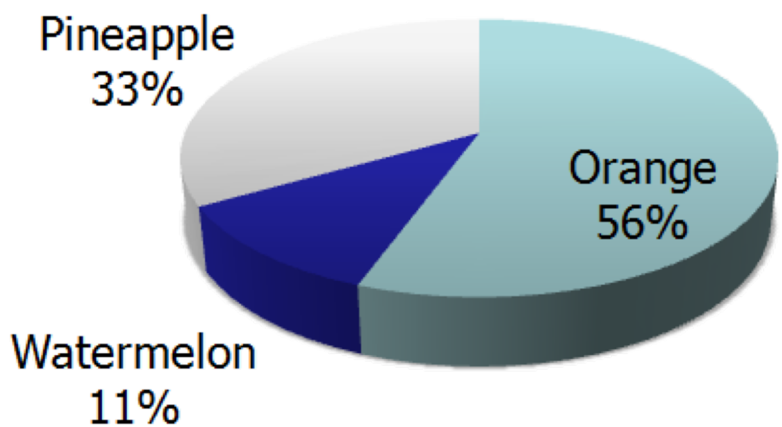



# Introduction to constrained designs

- D-optimal designs
- Mixture designs

# Case I - Fruit Punch

- A watermelon producer wants to introduce Watermelon juice into a fruit punch
- The fruit punch should be a blend of watermelon, orange and pineapple with at least 20% of watermelon
- Which combination of the 3 juices will yield highest consumer acceptance?



# Case II - Cooked Meat

- A ready meal manufacturer wants to develop a new meat-based meal.
- The cooking process involves both steaming and frying; each stage can vary between 5 and 15 minutes.
- Under 16 minutes total cooking time, the meat is still raw; over 24 minutes, overcooked. Such samples cannot be rated by the sensory panel.
- Is there a type of experimental design that can take into account those constraints?

# What Do These Cases Have In Common?

- These problems can be solved by using experimental design....
- But: the design variables involved cannot vary completely independently from the others
  - Case 1: the proportion of fruit juices must add up to 100%
  - Case 2: the total cooking time must not be under or above a certain limit
- **These are constrained situations**

# Factorial designs fail! (1/2)

- Cooked Meat example (case II)
  - Design variables: Marinating Time, Steaming Time, Frying Time
  - Responses: Sensory measurements

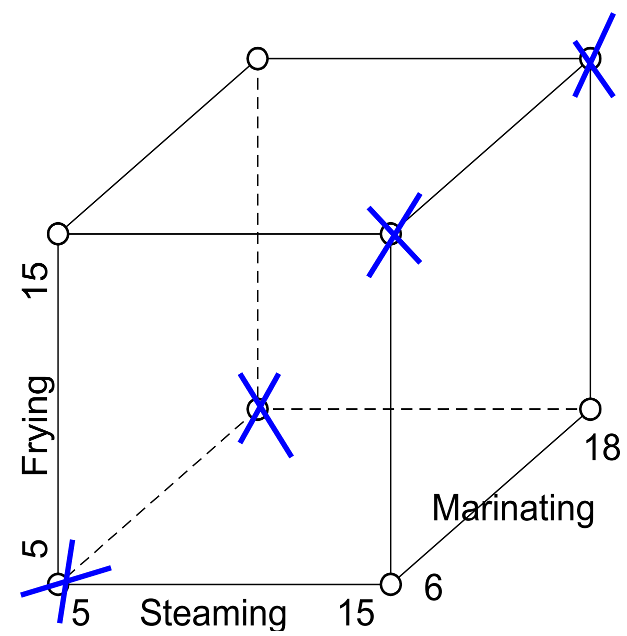
| Sample | Marinating | Steaming | Frying |
|---|---|---|---|
| 1 | 6 | 5 | 5 |
| 2 | 18 | 5 | 5 |
| 3 | 6 | 15 | 5 |
| 4 | 18 | 15 | 5 |
| 5 | 6 | 5 | 15 |
| 6 | 18 | 5 | 15 |
| 7 | 6 | 15 | 15 |
| 8 | 18 | 15 | 15 |

- Full Factorial solution
- 8 experiments combining the low and high levels of the variables

# Factorial designs fail! (2/2)

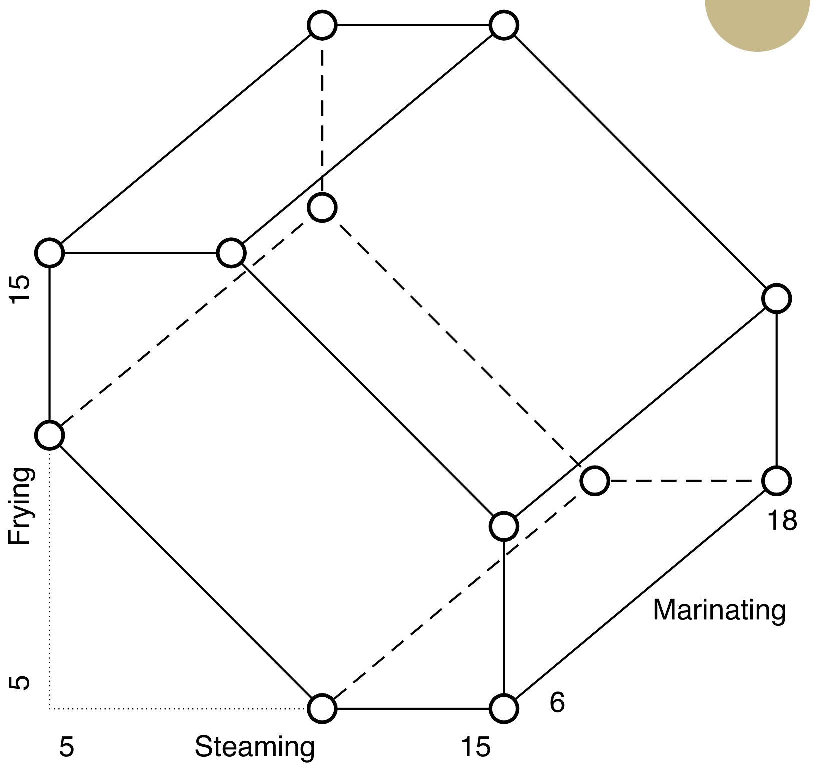
- Extreme combinations are forbidden
  - Steaming + Frying < 16 : raw meat
  - Steaming + Frying > 24 : overcooked
- In both cases sensory assessments are impossible

- Full Factorial does not apply
  - 4 out of 8 cube samples are excluded
  - The remaining 4 are not enough to explore the region of interest



# Constrained experimental region

- Multi-linear constraints:
  - Steaming + Frying ≥ 16 min
  - Steaming + Frying ≤ 24 min
- The experimental region becomes a polyhedron
- Orthogonal design no longer possible




There are several optimization criteria, the most common are:

- I-Optimality (gives the best distribution inside the design space)
- D-optimality (Maximizes the determinant of X; gives the smallest errors in the regression coefficients)

# What is a Mixture?

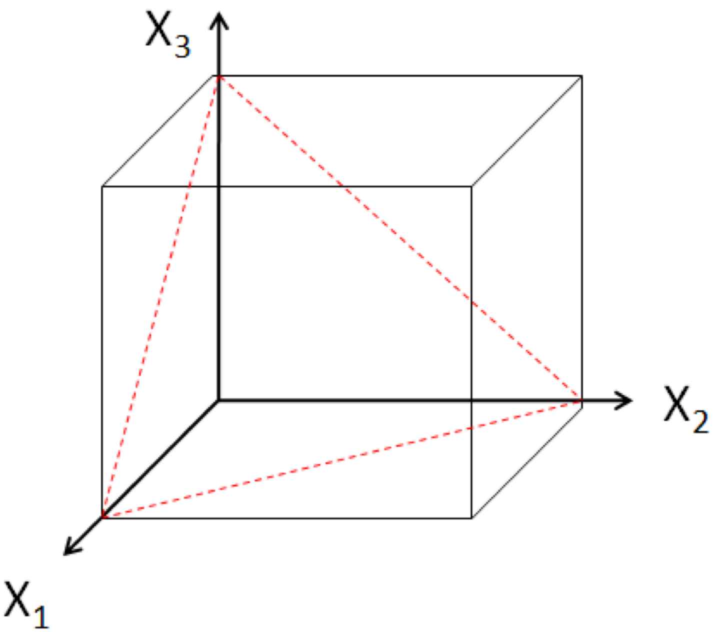
- The total quantity of ingredients is fixed
  - Example: A + B + C = 100%

- One dimension has “vanished”
  - Once you choose the amounts of A and B, the amount of C is fully determined

- The Mixture Problem
  - The measured response is assumed to depend only on the proportions of the ingredients and not the amount of the mixture

# Mixture Experimental Region

- 3 dimensions collapsed into 2
  - For a mixture of 3 ingredients the experimental region becomes flat
- This shape is called a simplex
- The simplex region contains all possible combinations of the 3 ingredients




Constraints:

$$\sum_{i=1}^{q} x_i = 1 \qquad x_i \geq 0$$

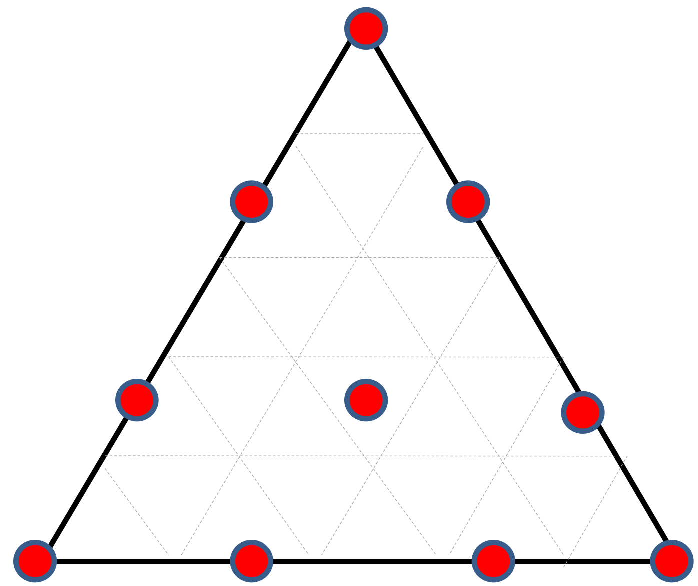
*q* = mixture components

# Types of Mixture Designs (1/2)

- The [q,m] Simplex Lattice Design
- The Simplex Lattice design has the form [q,m] where q is the number of mixture components and m is the order of the design to be supported.
- The design consists of the following mixtures:

  $\frac{0}{m}, \frac{1}{m}, \frac{2}{m}, \ldots, 1$

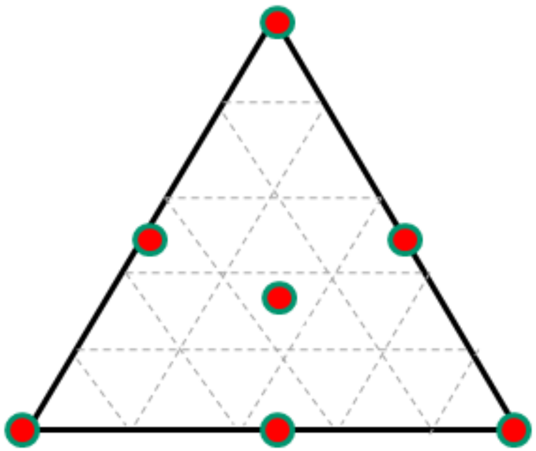
- Excellent Designs for investigating the extremes of a design space



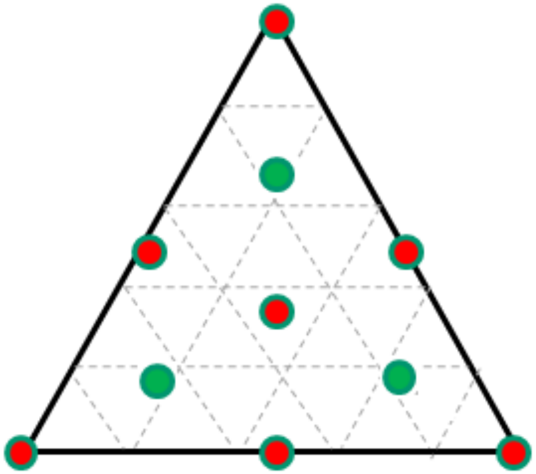
The [3,3] Simplex Lattice Design

# Types of Mixture Designs (2/2)

- The Simplex Centroid Design
- In a q component simplex, the number of distinct points is 2q-1
- Simple design for investigating the entire region economically.

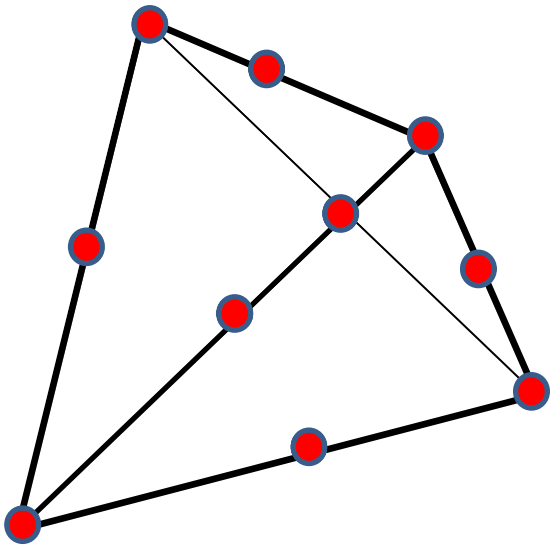


- May be further augmented with axial check blends



# Classical Mixture Designs

- The Simplex Lattice and the Simplex Centroid Designs can be analyzed using standard algorithms, as they fit precise mathematical models.
- It is easy to view these models up to 4 components.
- However, when certain constraints are imposed on a mixture, the simplex shape can be lost and alternative modelling procedures must be investigated.



[4,2] simplex lattice

# Analysis of mixture designs: Scheffe Models

- Mixture designs are generated to fit exact models
  - First order:

$$Y = \sum_{i=1}^{q} \beta_i X_i$$

  - Second order:

$$Y = \sum_{i=1}^{q} \beta_i X_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^{q} \beta_{ij} X_i X_j$$

  - Special cubic:

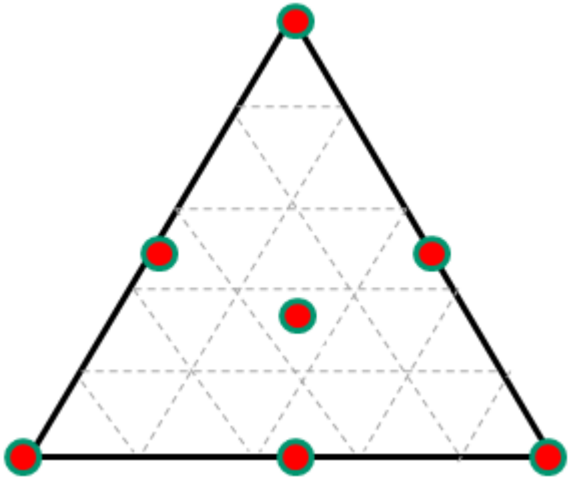
$$Y = \sum_{i=1}^{q} \beta_i X_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^{q} \beta_{ij} X_i X_j + \sum \sum_{i<j<k} \sum^{q} \beta_{ijk} X_i X_j X_k$$

# Example: Simplex Centroid Design with Three Components

- q = 3:
  - Require all mixtures of the form 1/q (i.e. 1, 1/2 and 1/3)
- Fit the special cubic model

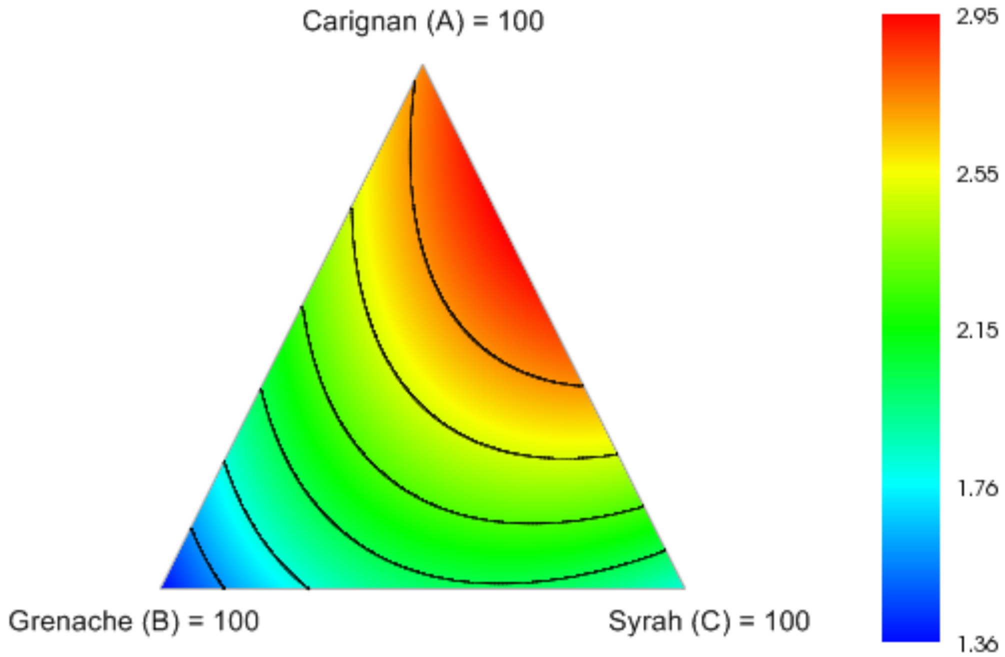
$$Y = \sum_{i=1}^{3} \beta_i X_i + \sum_{i=1}^{2} \sum_{j=2}^{3} \beta_{ij} X_i X_j + \sum \sum_{i<j<k} \sum^{3} \beta_{ijk} X_i X_j X_k$$

- This results in 7 terms in the equation corresponding to the 7 points in the design



# Simplex Response Surface

- 2D-map of three mixture components
- The three coordinates sum to 100%
- Point the mouse on the surface to get optimal mixture



# Summary (1/2)

- DoE is the best way of generating meaningful experiments that will provide the **maximum information** in the **minimal experimental effort**
- Designed experiments can be performed sequentially, i.e. **more information can be added** if need be, to an existing design
- Fractional factorials can be **simplified** when factors are found to be insignificant, resulting in more precise results, without any further experimentation being done
- Many factors can be analyzed in a **small number of experiments** to screen out important factors

# Summary (2/2)

- Remember to check the **residuals** of the model to see if the assumptions of the ANOVA hold or to detect outliers

- When a small number of factors have been isolated, the design can be **extended** to become an optimization design

- When dealing with mixtures, the experimental designs become **constrained**. A special class of designs must be considered to analyze these.

# Data for exercise

- https://www.dropbox.com/s/gh9kmhqcb8krskq/SmallDemoFactorialDesignFrank.unsb?dl=0
- https://www.dropbox.com/s/u9nypa2pkhkrtyw/GlutenStarchPowdersNIR.unsb?dl=0