Experimental design

Roger PHAN-TAN-LUU

Université Paul Cézanne - FRANCE

CEMRACS 2006 – CRIM-Marseille - July

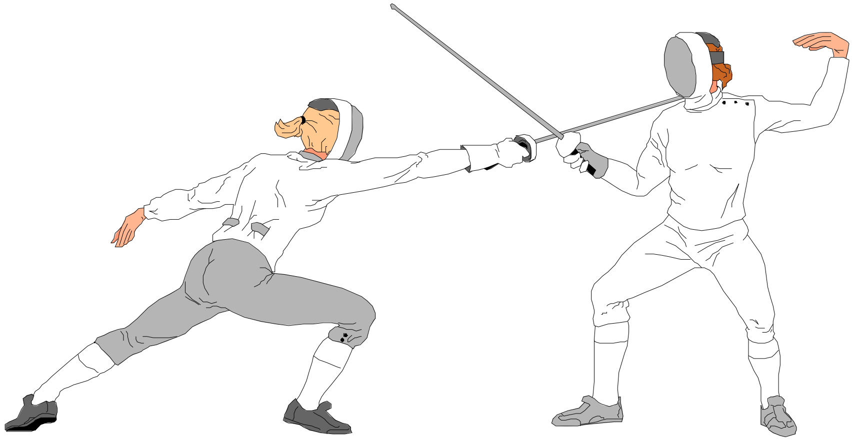
Experimental design

Experimentation

- Fisher R. (1917)
- Box G.E.P. et Wilson (1954)
- Tagushi G. (1960)
- Kiefer J. C. (1958)
- Mitchell T. (1968)
- Fedorov V. V. (1972)
- …………..

Simulation

- Klejnen J.P.C. …… (1970),
- McKay M.D.. ………..(1979),
- Morris M.D………….(1995),
- Sacks J. ………………(1989),
- …………..



Duel ???

Why ?

	Experimentation	Simulation
Number of factors	Small or moderate (<20)	Large 30, 100, 1000, …
Model	Black box model	Complex model
Forme du modèle	Linear (coefficients)	Non-linear, non-polynomial,..
Effets d'interaction	low-order (< 3)	Substantial higher-orders
Errors	homogeneous	heterogeneous
Domain	symmetric	constraints
………………………………	………………………………………	…………………………………………

Simulation ? ***Experimentation ?***

- Performance measures

- Responses

Response

The response is a result :

we cannot directly act on it

OUTPUT VARIABLE

~~Properties:~~

~~- *pertinent,*~~
~~- *reproducible,*~~
~~- *known with an acceptable precision,*~~

How can we modify the response ?

We must act indirectly !

Factors

We can directly act on the value of the factor

INPUT VARIABLES

Experimental domain of the factors

Each factor can be set to two or more values, called ***factor levels***

A factor can be either :

qualitative **quantitative**

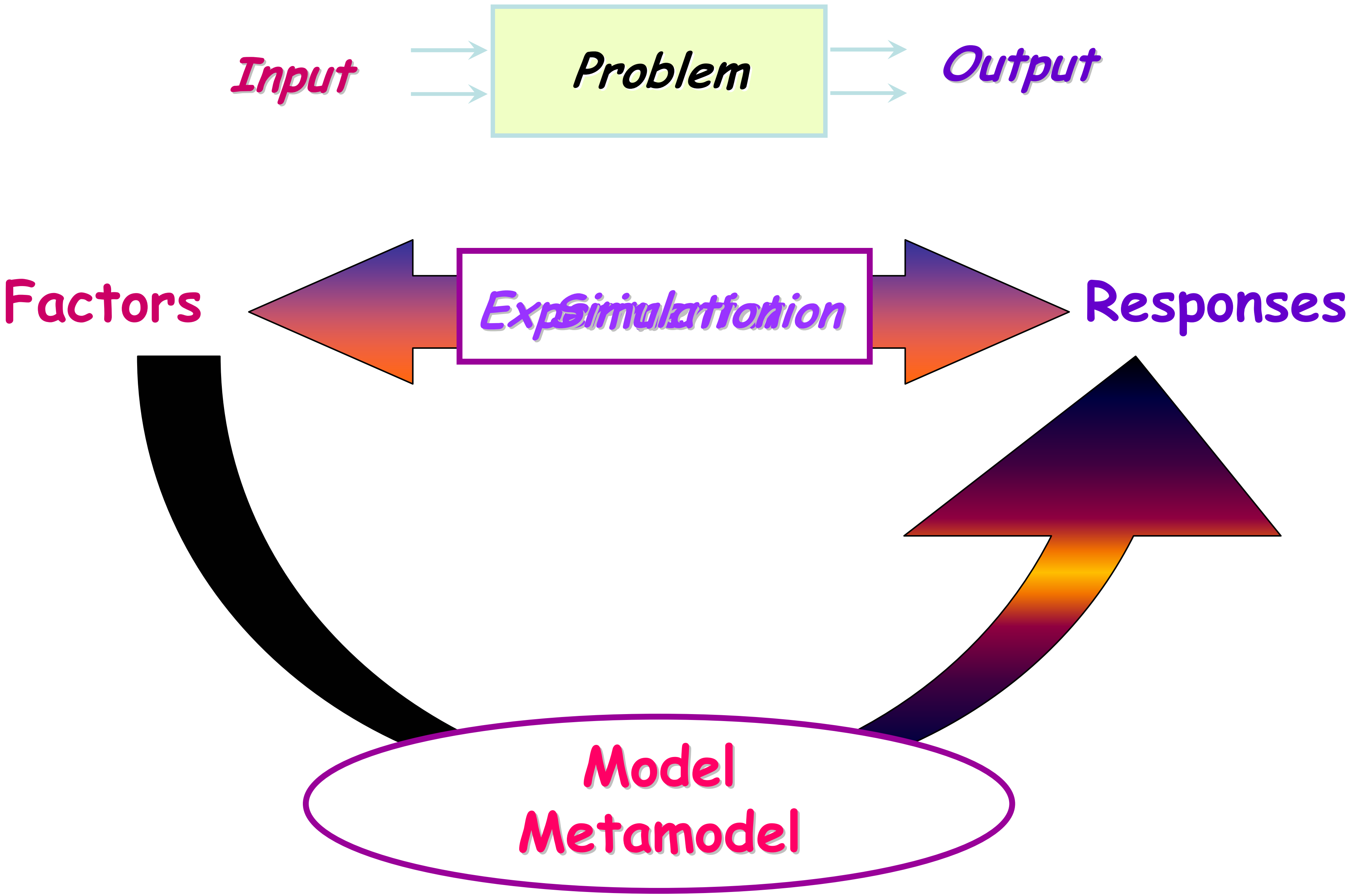
The various levels that a factor can have

↓

Experimental domain of the factors

↓

Experimental domain of interest


Input
Problem
Output
Factors
Experimentation
Simulation
Responses
Model
Metamodel

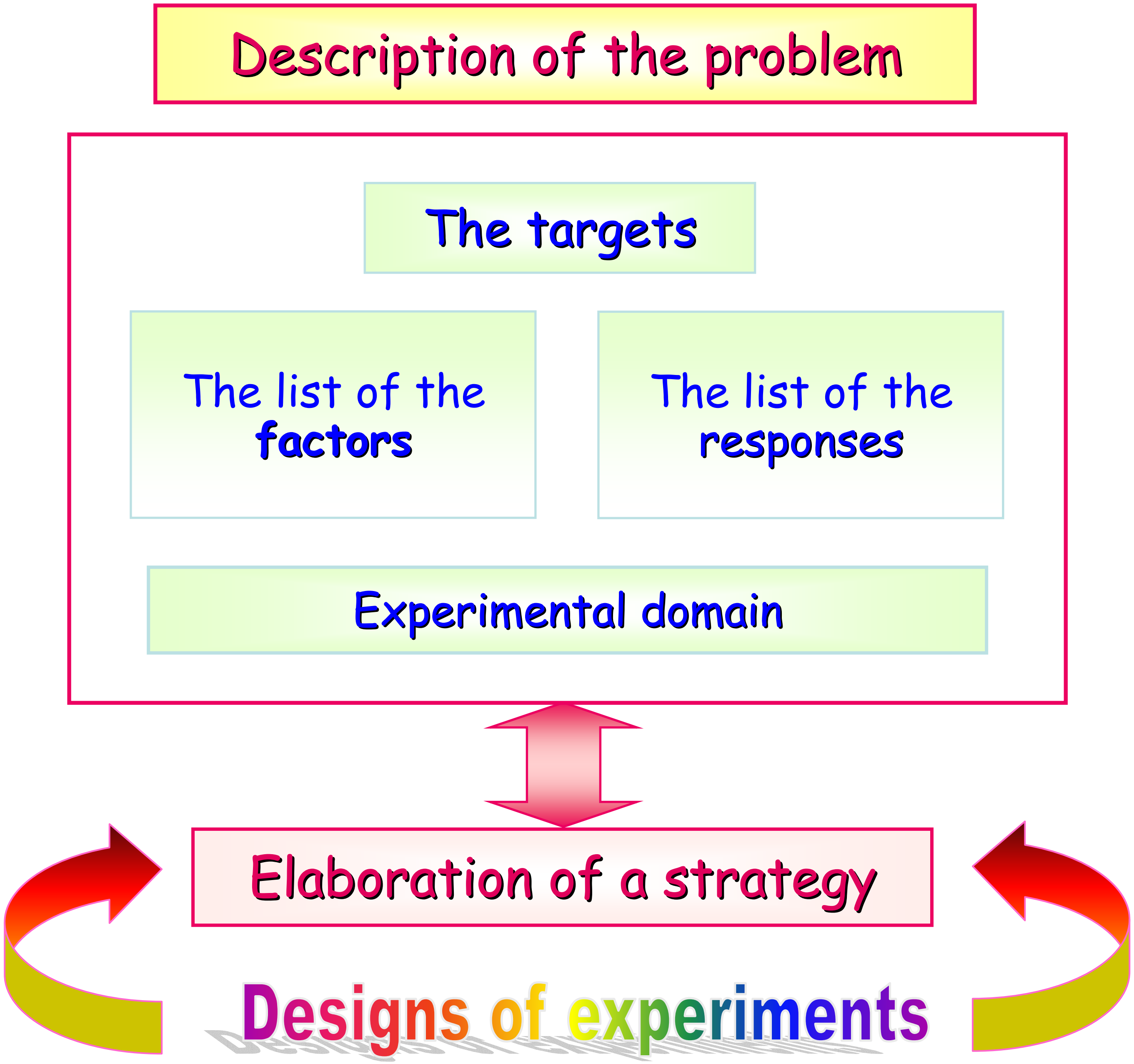
We must do experiments (simulations)

Which ones ?

Those, bringing the desired information !

Experimental strategy

D.O.E.


Description of the problem
The targets
The list of the factors
The list of the responses
Experimental domain
Elaboration of a strategy
Designs of experiments

To elaborate the experimental strategy

to choose an appropriate Experimental design in accordance with

THE TARGETS

Exploratory research

Screening of factors

Quantitative study of factors

Optimization

Design of experiments

To elaborate the experimental strategy

➡ *to choose an appropriate Experimental design in accordance with*

THE TARGETS

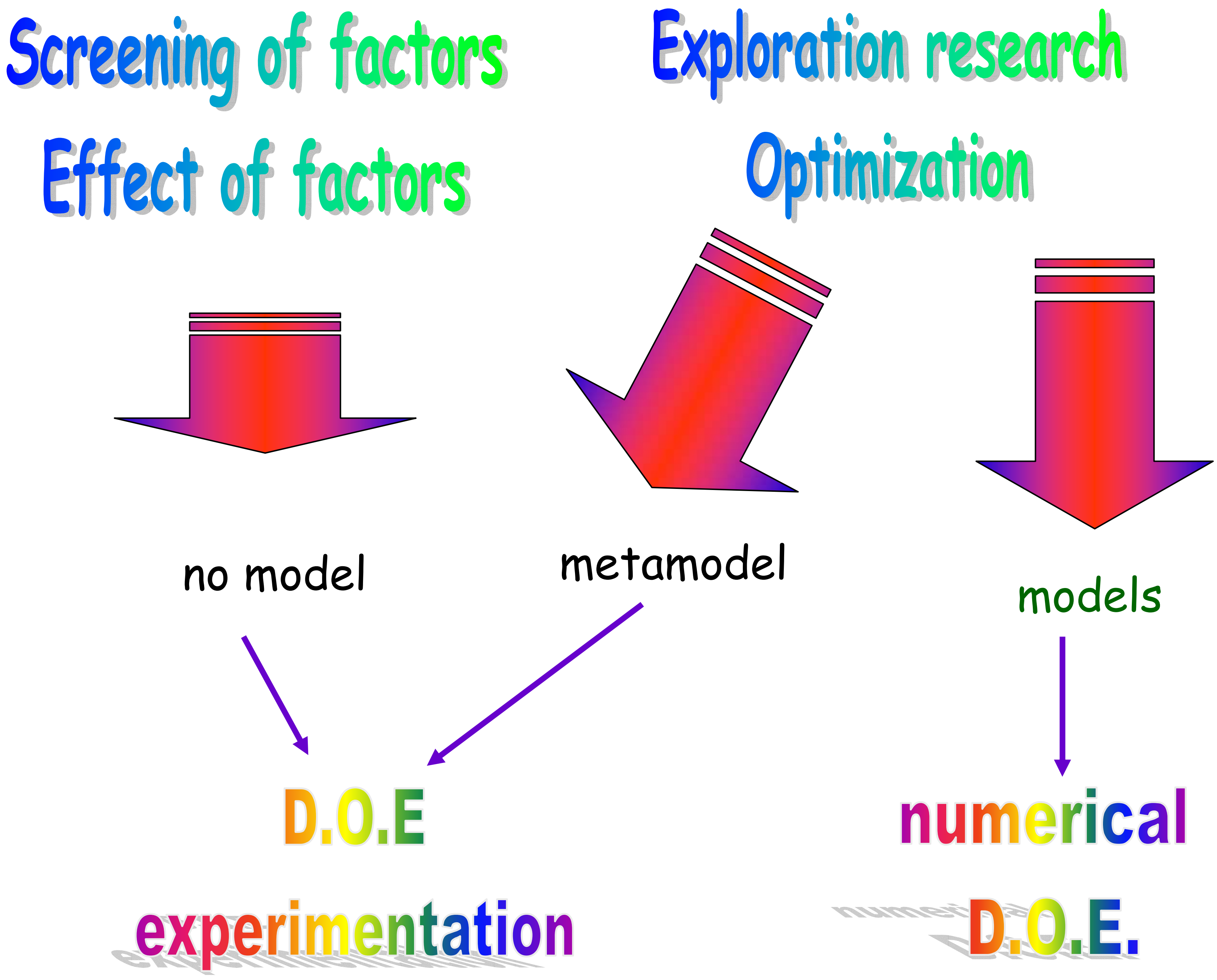
- to determine the operability region (range of values for each factor in which the system can operate),
- to investigate the domain

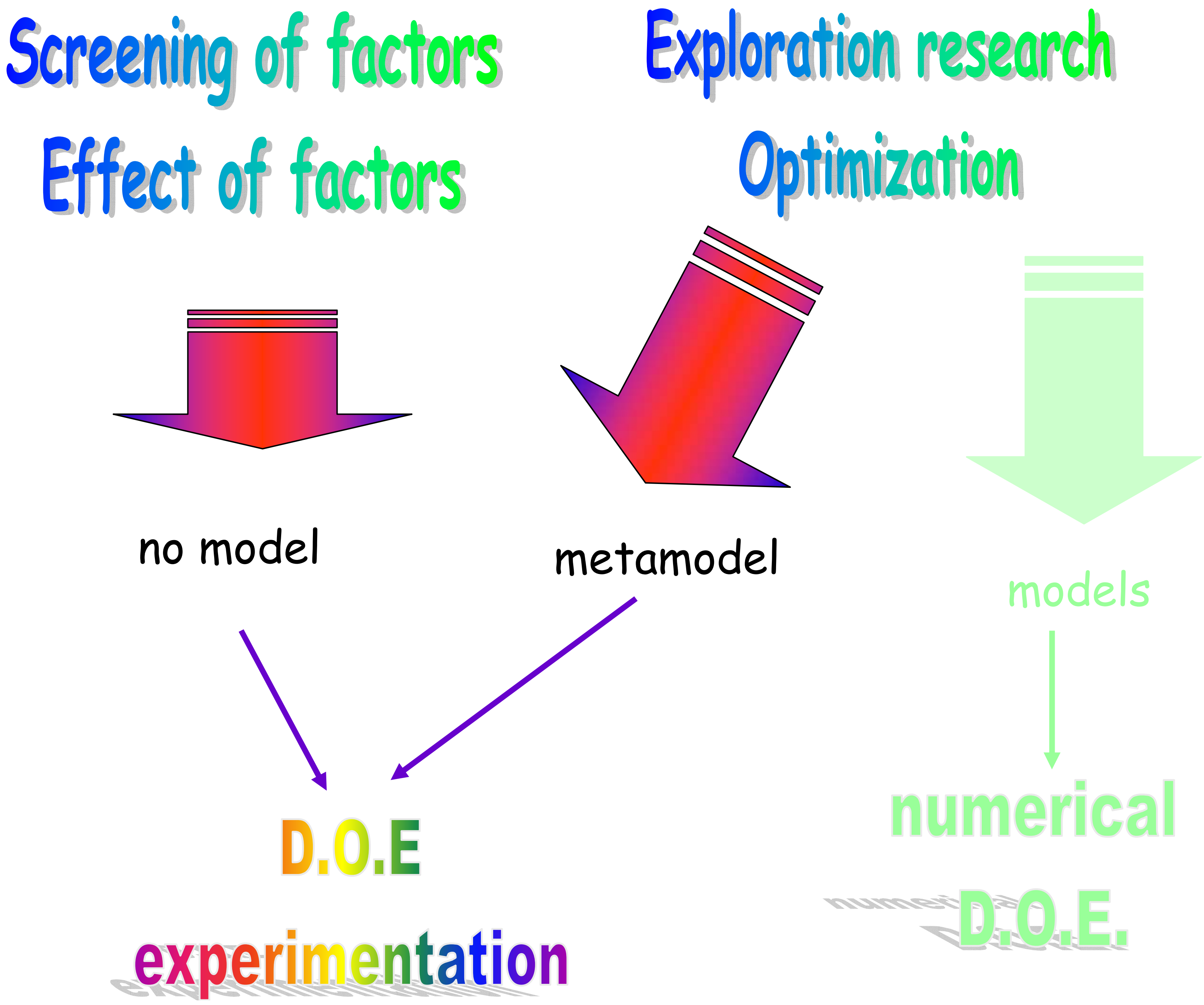
To investigate quickly, among a set of potentially influential factors, those which are really influential.

To quantify the effects of the factors and the interaction effects

To know, anywhere in the experimental domain, the value of the respones

Design of experiments


Screening of factors
Effect of factors
Exploration research
Optimization
no model
metamodel
models
D.O.E
experimentation
numerical
D.O.E.


Screening of factors
Effect of factors
Exploration research
Optimization
no model
metamodel
models
D.O.E
experimentation
numerical
D.O.E.

To elaborate the experimental strategy

to choose an appropriate Experimental design in accordance with

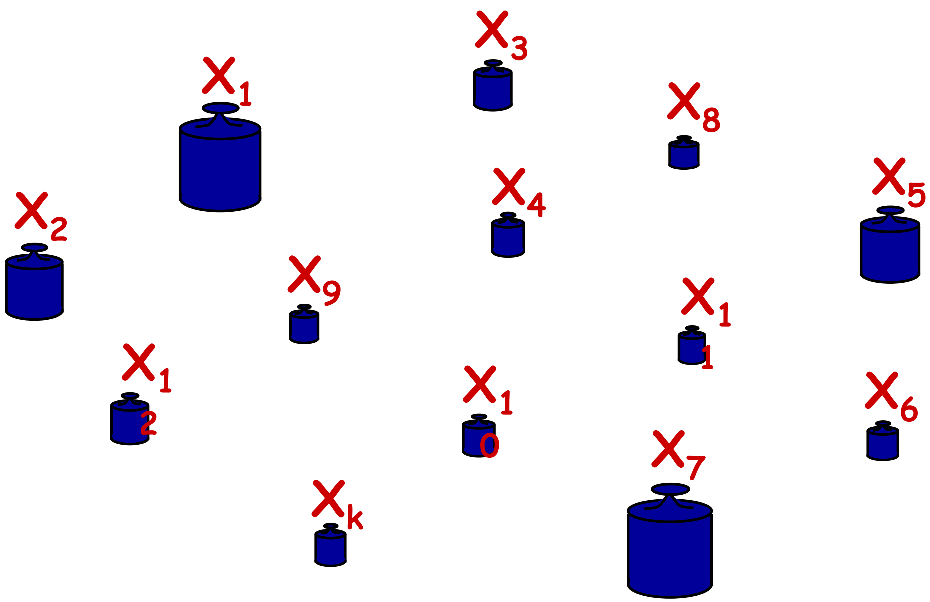
THE TARGETS

Screening of factors

Screening of factors

➡ *searching for the few **really important factors**, among many potentially influential factors.*

To know the "weight" of the factors



Screening of factors

Strategy allowing to identify really important factors (h) among a lot of potentially important factors (k)

$$h << k$$

HYPOTHESIS : Additivity

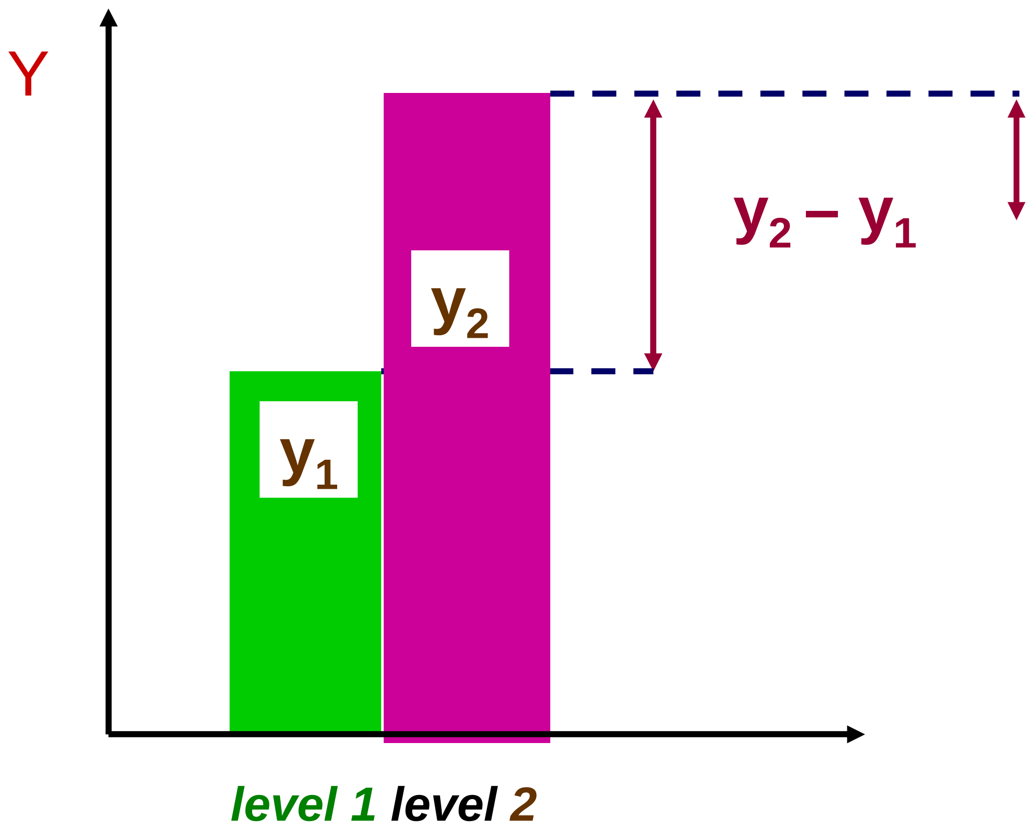
1 kg apple + 1 kg orange marmelade = 2 kg

1 kg pear + 1 kg apricot jam = 2 kg

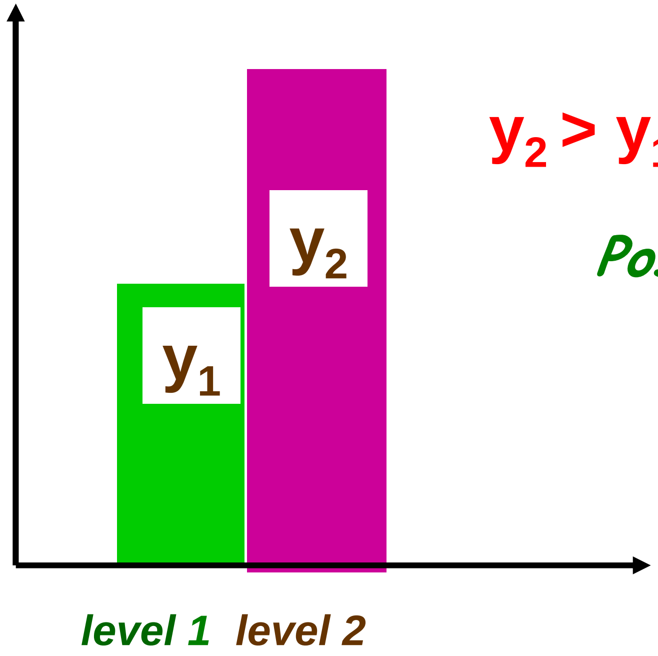
N°	X_1	Y
1	- 1	y_1
2	1	y_2

$$y_1 = b_0 - b_1$$
$$y_2 = b_0 + b_1$$

$$b_0 = (+\ y_1 + y_2)\ /\ 2$$
$$b_1 = (-\ y_1 + y_2)\ /\ 2$$

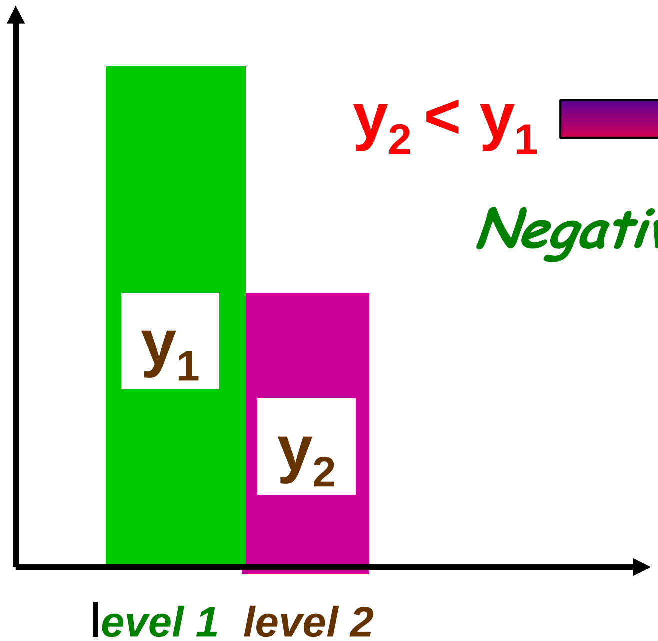


$$(y_2 - y_1)\ /\ 2 = b_1$$



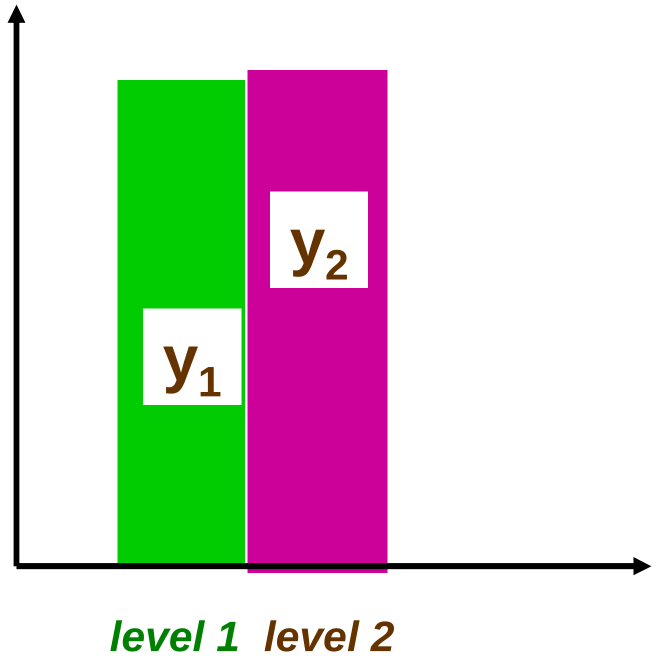
$y_2 > y_1 \Rightarrow b_1 > 0$

Positive weight



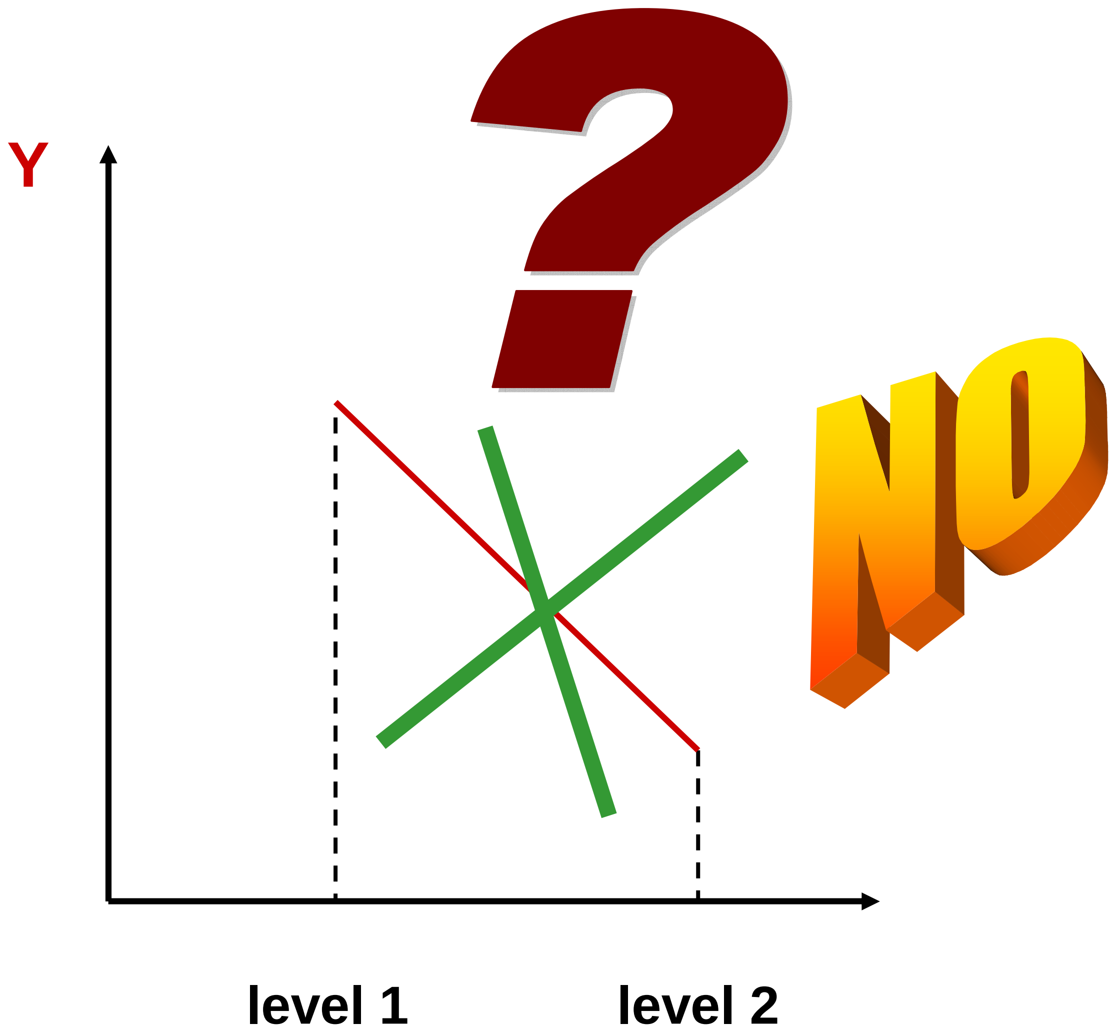
$y_2 < y_1 \Rightarrow b_1 < 0$

Negative weight

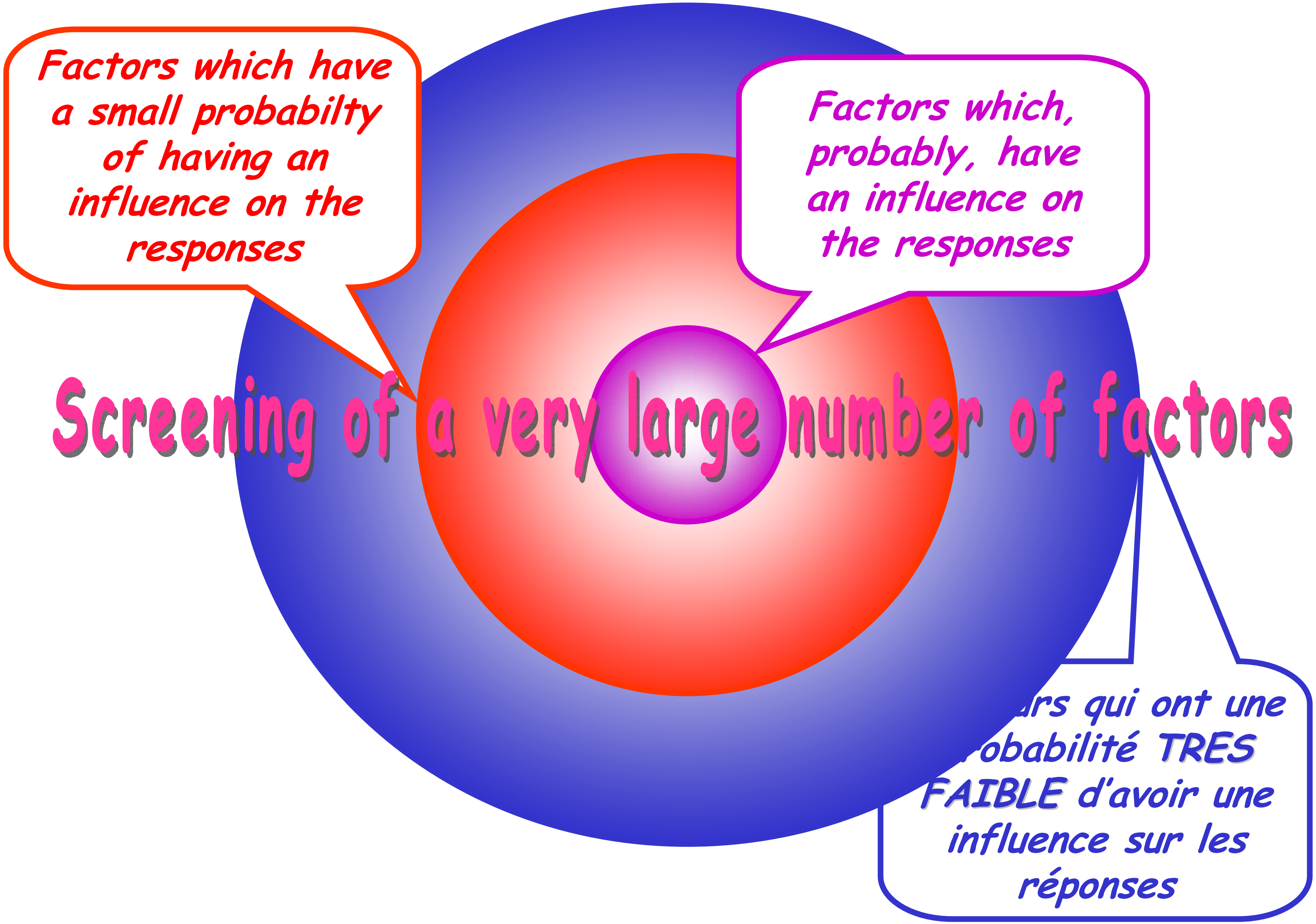


$y_2 \approx y_1 \Rightarrow b_1 \approx 0$

Weight nul



Y
NO
level 1
level 2



Factors which have a small probabilty of having an influence on the responses
Factors which, probably, have an influence on the responses
Screening of a very large number of factors
urs qui ont une
obabilité TRES
FAIBLE d'avoir une
influence sur les
réponses

The probability that a factor is active is VERY SMALL

The number of active factors is very small !

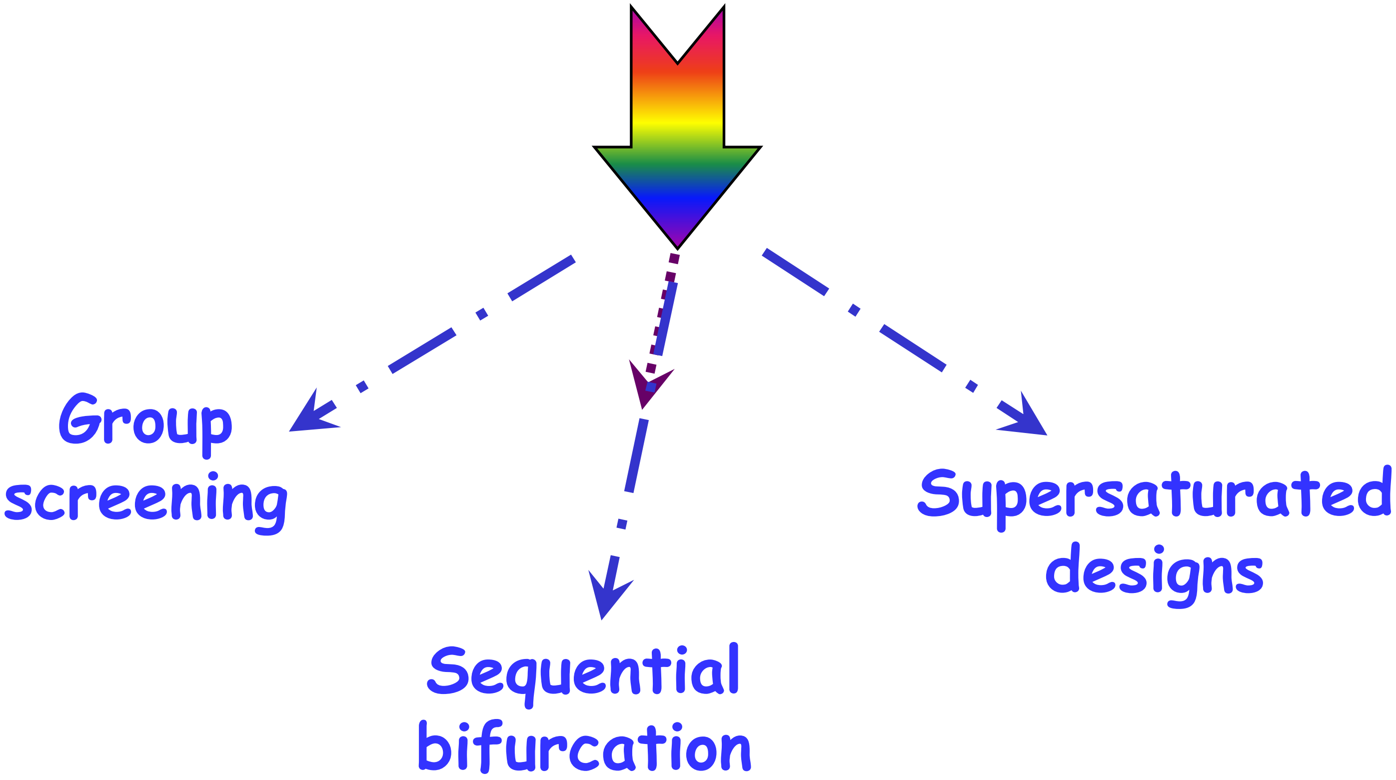
Problem in a production

Robustness of an analytical method

Simulation models

The probability that a factor is active is VERY SMALL

To quickly identify the few factors, which are really influential



Group screening, Sequential bifurcation

Hypothesis :

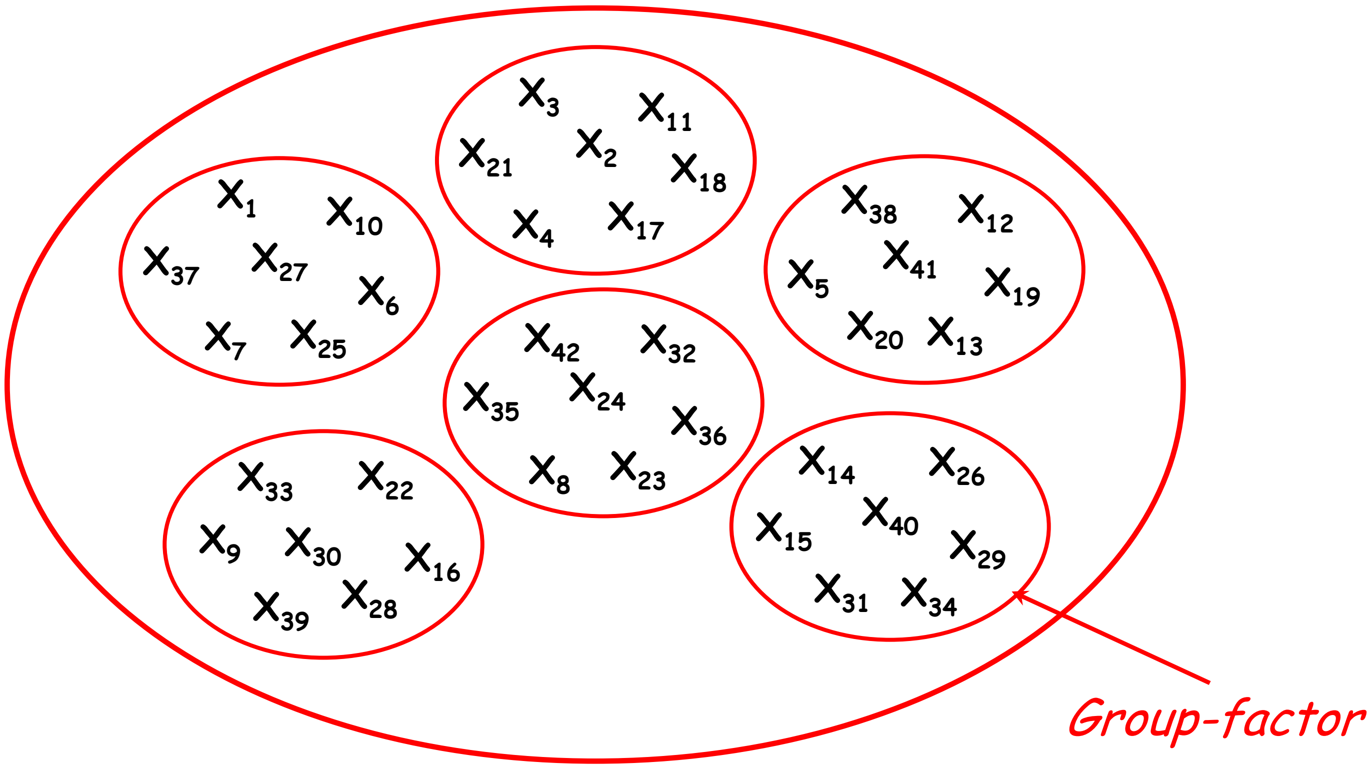
- Only a few factors are really important
  "parsimony principle" or " effect sparsity "
- There is no interaction effect between factors
- Each factor has 2 levels and we know the signs or directions for the possible effects
- The effect of a group-factor is significant if and only if the effect of at least one of its factors is significant
- If a group is not influential, then all the factors in this group are eliminated

Group screening

WATSON G. S. (1961)

➡ The k factors are grouped in several groups,

Step1 :

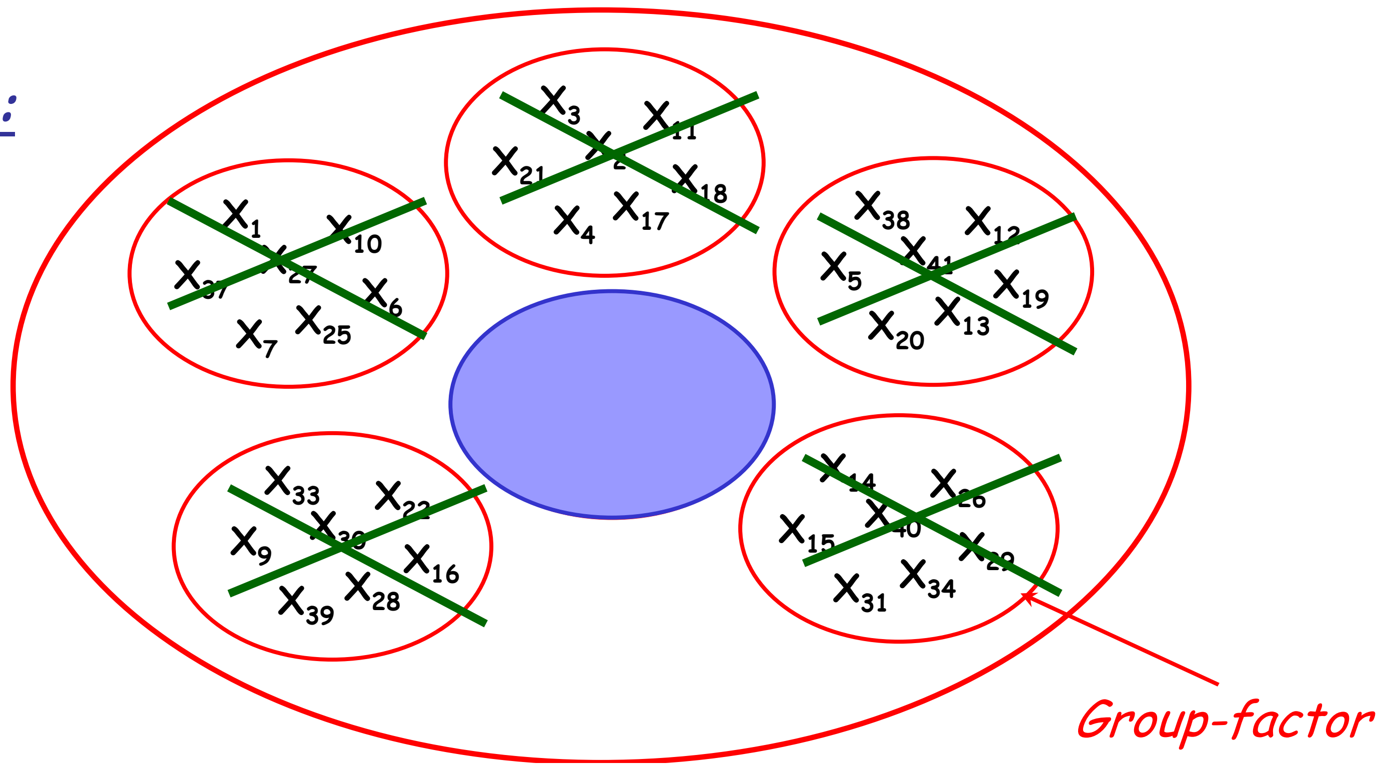


Group screening

WATSON G. S. (1961)

- The k factors are grouped in several groups,
- In the 1[st] step, each group is treated as a "factor" (group-factor) and the "non active" groups are eliminated,

Step 2 :



Group screening

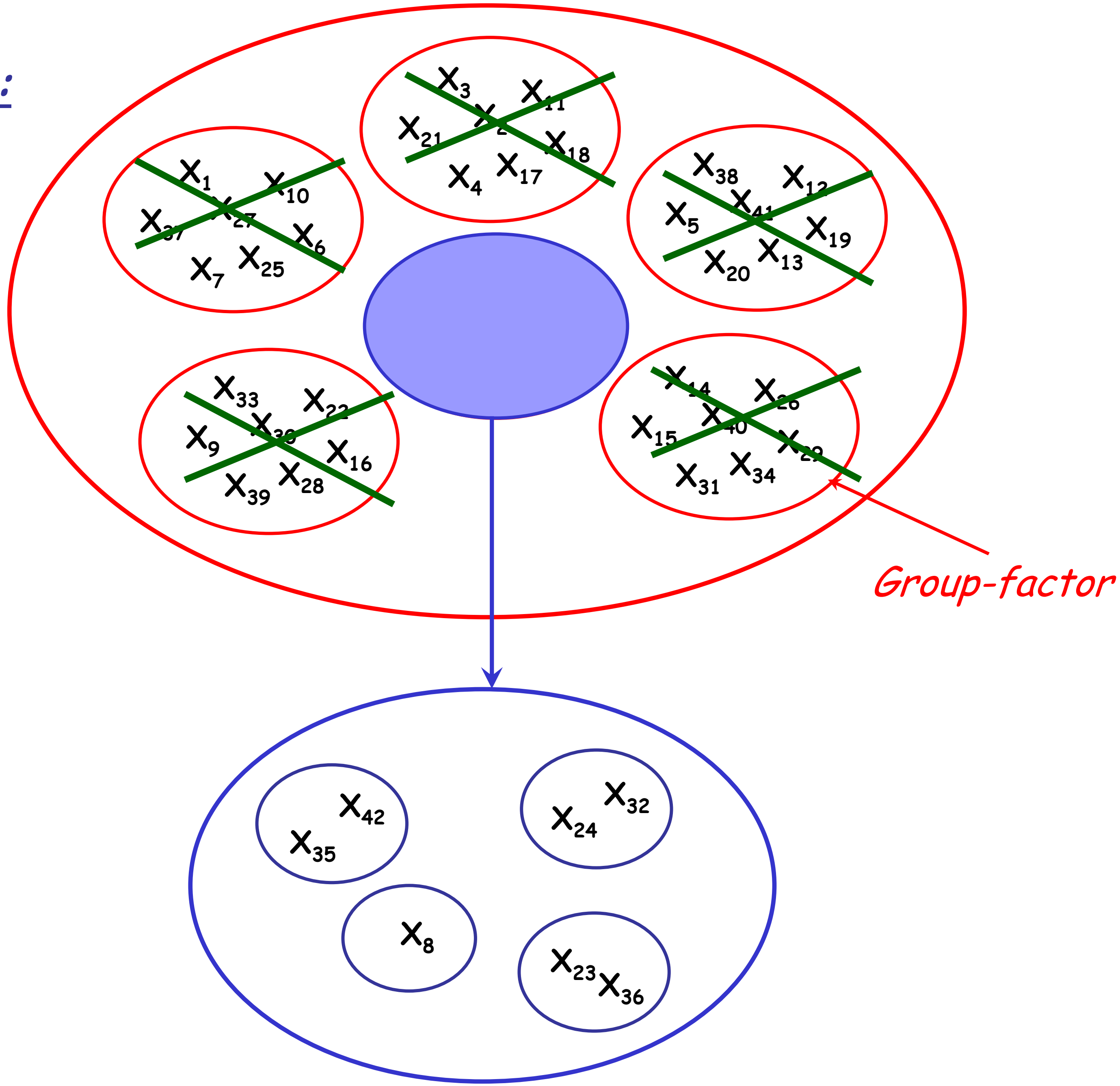
WATSON G. S. (1961)

- The k factors are grouped in several groups,

- In the 1st step, each group is treated as a "factor" (group-factor) and the "non active" groups are eliminated,

- The factors in the "active groups", will be studied in a subsequent experiment,
  - individually (process in 2 steps)
  - in smaller groups (process in several steps)

Step 2 :



X_{1} X_{10} X_{37} X_{27} X_{6} X_{7} X_{25}
X_{3} X_{11} X_{2} X_{21} X_{18} X_{4} X_{17}
X_{38} X_{12} X_{41} X_{5} X_{19} X_{20} X_{13}
X_{33} X_{22} X_{30} X_{9} X_{16} X_{39} X_{28}
X_{14} X_{26} X_{40} X_{15} X_{29} X_{31} X_{34}
Group-factor
X_{42} X_{35}
X_{32} X_{24}
X_{8}
X_{23} X_{36}

Step 1 :

k factors → g_1 groups of f_1 factors

$$f_1 = k / g_1$$

Group-factor
- level (+) → all factors : (+)
- level (-) → all factors : (-)

↓

Design of experiments

(g_1+1) experiments, at least

Experimentation → r_1 active groups → $r_1 f_1$ factors

Step 2 :

$r_1 f_1$ factors → g_2 groups of f_2 factors

$$f_2 = r_1 f_1 / g_2$$

Experimentation → r_2 active groups → $r_2 f_2$ factors

....

Step C :

$r_{c-1} f_{c-1}$ factors → g_c groups of f_c (=1) factors

↓

Experimentation → r_c active factors

After the theory, …….

1 – A factor whose the direction of the effect is not well known, must be placed ALONE in a group

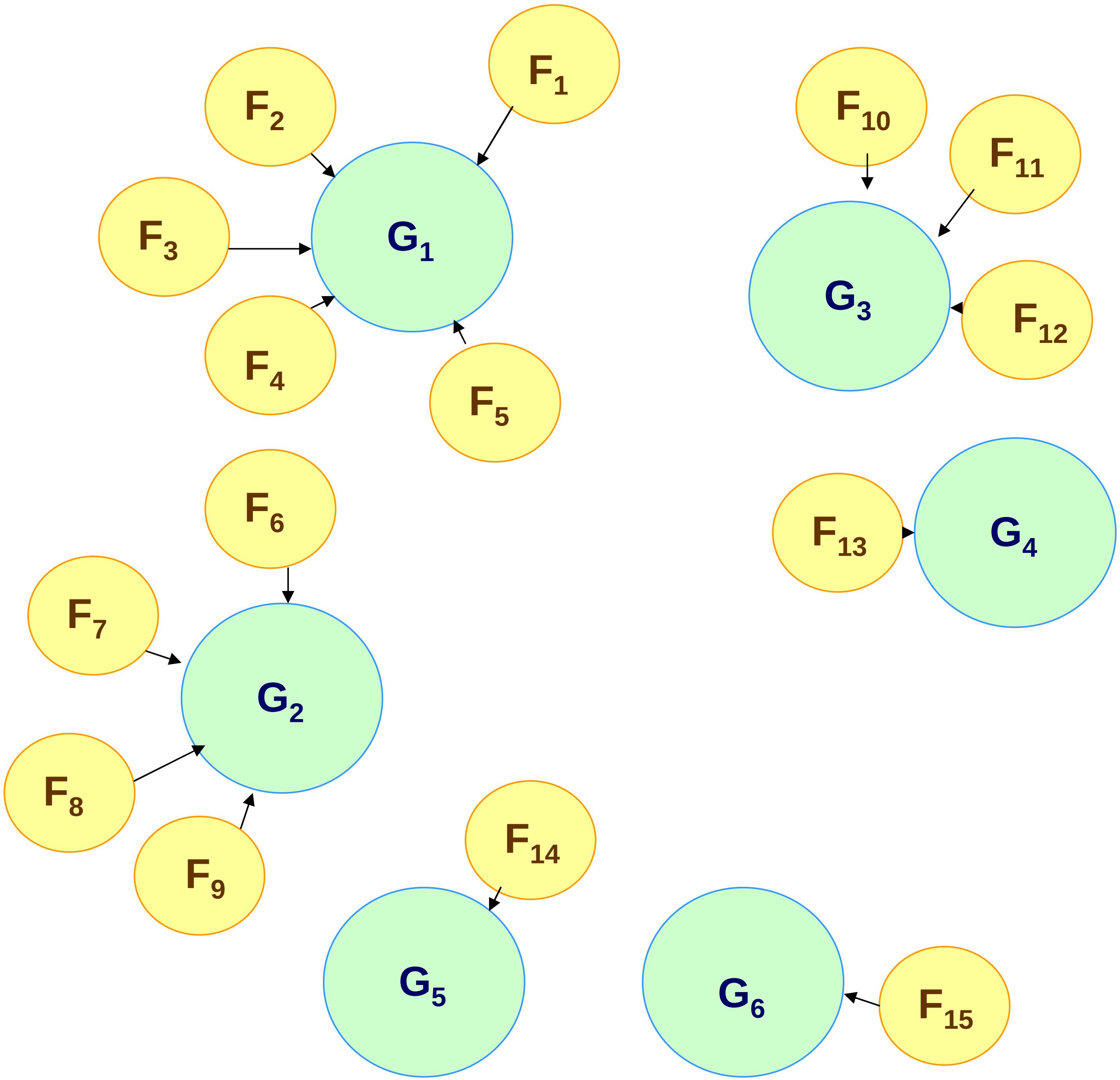
2 – The factors which are supposed as having an important positive effect are put TOGETHER in a group

3 – The factors which are supposed as having a small positive effect are put TOGETHER in a group

4 – The factors which are supposed as having an possible positive effect are put TOGETHER in a group

Example : k = 15 factors

- 5 factors F1, .., F5 are supposed as having an important positive effect. They are put TOGETHER in the group G1
- 4 factors F6, .., F9 are supposed as having an possible positive effect. They are put TOGETHER in the group G2
- 3 factors F10, F11, F12 are supposed as having a small positive effect. They are put TOGETHER in the group G3
- 3 factors F13, F14, F15 whose the direction of the effect is not well known, must be placed, each, ALONE in a group G4, G5, G6



STEP 1

Example : **Robustness of an analytical method**

Robustness tests

↓

To test the susceptibility of an analytical procedure to small changes in the experimental conditions

↓

1st step : the selection of the factors and levels to be tested

2nd step : the selection of the most suitable experimental design

only a small number of significant effects is expected

Group screening design

Example : **Robustness of a pharmaceutical development**

Development of a new drug

Development of a new active ingredient

The process must be "robust" !

Example : Robustness of a pharmaceutical development

k = 35 factors

FACTORS	Range
1. Charge pyridine	± 5 %
2. Addition temperature	± 3°C
3. Reaction temperature	± 3°C
4. Reaction time	90 to 120 min
5. Addition H_2O	120 to 140 min
6. Charge acetone	± 5 %
....	....
35. Drying	45 ± 5°C

Example : **Robustness of a pharmaceutical development**

k = 35 factors

Which are the influential factors ?

Step 1 : → **7 groups** (5 factors by group)

Plackett and Burmann design N = 8

y

1 active group → **5 factors**

P. B. design

N = 8

2 active factors !

In all :

N = 8 + 8 = 16 exp.

Sequential Bifurcation

Kleijnen J. P. C. and Bettonvil B

➡ *searching for the few **really important factors,** among many potentially influential factors.*

⇓

"dichotomic" approach

⇓

Effective and efficient method !

Sequential Bifurcation

Kleijnen J. P. C. and Bettonvil B

y [k] : experimental response when all factors are at their high level (+)

y [0] : experimental response when all factors are at their low level (-)

y [j] : experimental response when :

- the factors 1, ..., j are at their high level (+)
- the remaining factors j+1, ..., k are at their low level (-).

If y [0] = y [k] ⇒ no influential factor

Otherwise ⇒ some factors are influential

⇒ y [k/2] : k/2 factors (+) et k/2 factors (-)

Example : Robustness of a pharmaceutical development

k = 35

Hypothesis : 2 active factors : X_8, X_{28}

$y(0)$
$y(35)$

$y(0) < y(35)$

$\beta_{1-35} > 0$

$y(18)$

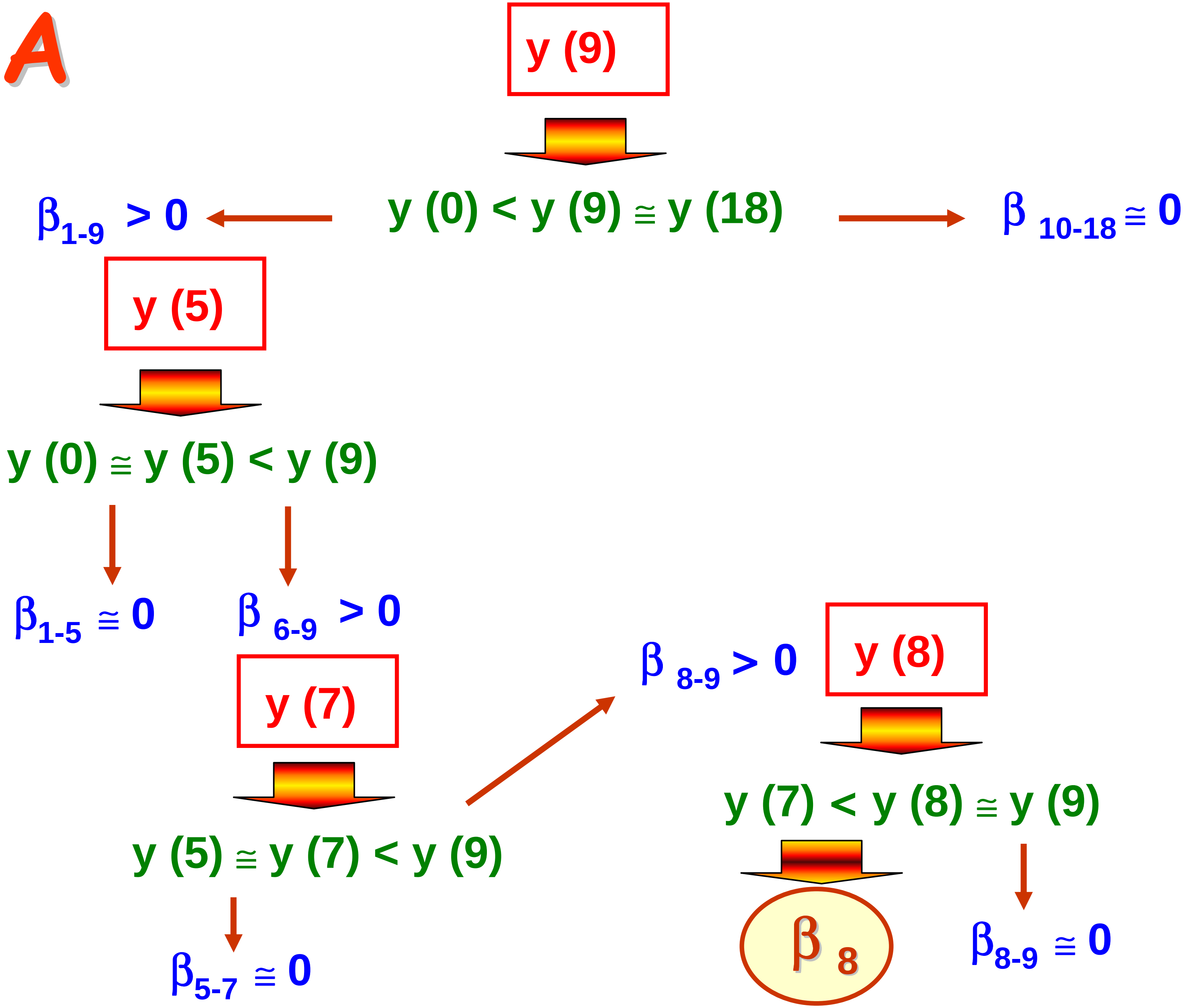
$\beta_{1-18} > 0$

$y(0) < y(18) < y(35)$

$\beta_{19-35} > 0$

A

B


A
y (9)
y (0) < y (9) ≅ y (18)
β 1-9 > 0
β 10-18 ≅ 0
y (5)
y (0) ≅ y (5) < y (9)
β 1-5 ≅ 0
β 6-9 > 0
y (7)
y (5) ≅ y (7) < y (9)
β 5-7 ≅ 0
β 8-9 > 0
y (8)
y (7) < y (8) ≅ y (9)
β 8
β 8-9 ≅ 0

B

y (26)

$y(18) \cong y(26) < y(35)$

$\beta_{18\text{-}26} \cong 0$

$\beta_{27\text{-}35} > 0$

y (31)

$y(26) < y(31) \cong y(35)$

$\beta_{26\text{-}31} > 0$

$\beta_{32\text{-}35} \cong 0$

11 experiments !

y (28)

$y(26) < y(28) \cong y(31)$

$\beta_{29\text{-}31} \cong 0$

y (27)

$y(26) \cong y(27) < y(28)$

$\beta_{26\text{-}27} \cong 0$

β_{28}

Sequential Bifurcation
with interactions

Kleijnen J. P. C. and Bettonvil B

y-[j] : *mirror* observation of y [j]

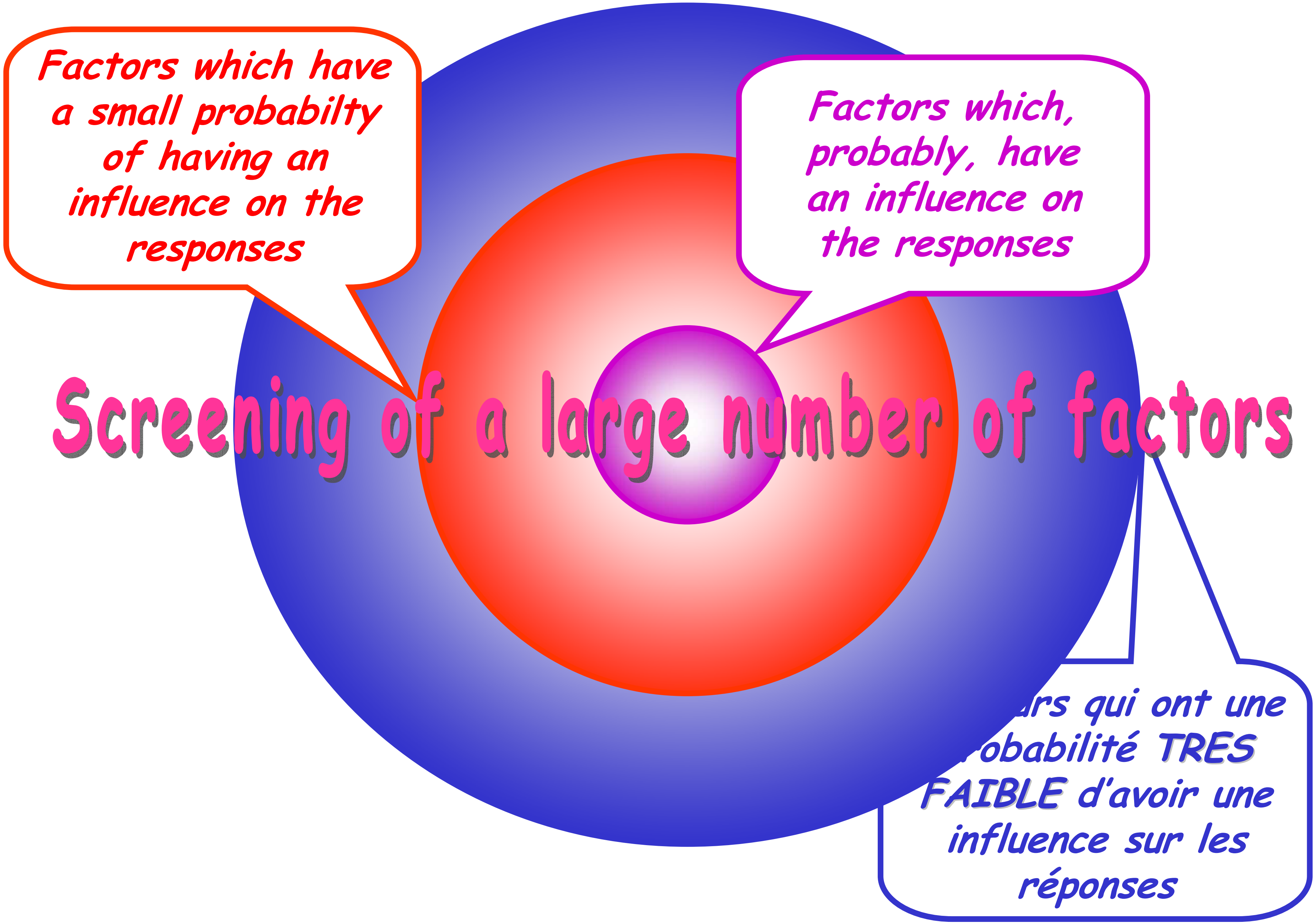
y [j] : experimental response when :

- the factors 1, ..., j are at their low level (-)
- the remaining factors j+1, ..., k are at their high level (+).

y-[0] = y [k]

y [k] = y -[0]

decision ⇒ (y[j] - y-[j]) - (y[t] - y-[t])


Factors which have a small probabilty of having an influence on the responses
Factors which, probably, have an influence on the responses
Screening of a large number of factors
urs qui ont une
obabilité TRES
FAIBLE d'avoir une
influence sur les
réponses

The probability that a factor is active is SMALL

k factors with 2 levels ($N \leq 100$)

Plackett-Burman designs

$N \geq k + 1$

$N \equiv 0$ (modulo 4)

Number of factors

Number of experiments

Number of levels

$2^k // N$

Plackett-Burman designs N = 0 (modulo 4)

PLACKETT R.L. et BURMAN J.P.

Design of optimum multifactorial experiments.

Biometrika, 1946, 33, 305.

N = 4 + + -

N = 8 + + + - + - -

N = 12 + + - + + + - - - + -

N = 16 + + + + - + - + + - - + - - -

N = 20 + + - - + + + + - + - + - - - - + + -

N = 24 + + + + + - + - + + - - + + - - + - + - - - -

....

N = 99 ....

Alias matrix

Postulated model $: \eta = X\beta$

« True » model $: \eta = X\beta + X_1\beta_1$

⇒ $B = (X'X)^{-1} X'Y$ $\quad E[B] = ?$

$$E[B] = E[(X'X)^{-1} X'Y] = E[(X'X)^{-1} X'\{X\beta + X_1\beta_1 + \varepsilon\}]$$

$$= E[(X'X)^{-1} X'X\beta + (X'X)^{-1} X'X_1\beta_1 + (X'X)^{-1} X'\varepsilon\}]$$

$$= E[\beta + (X'X)^{-1} X'X_1\beta_1 + (X'X)^{-1} X'\varepsilon]$$

$$= \beta + E[(X'X)^{-1} X'X_1\beta_1] + E[(X'X)^{-1} X'\varepsilon]$$

$$= \beta + E[(X'X)^{-1} X'X_1\beta_1] + (X'X)^{-1} X' E[\varepsilon]$$

$$= \beta + (X'X)^{-1} X'X_1 \beta_1$$

$$= \beta + A\beta_1 \qquad A = (X'X)^{-1} X'X_1$$

$$E[B] = \beta + A\beta_1$$

A : alias matrix

Postulated model

$$\eta = \beta X = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

ξ_4

N	X_1	X_2	X_3
1	-	-	+
2	+	-	-
3	-	+	-
4	+	+	+

« true » model

$$\eta = \beta X + \beta_1 X_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3$$

N	X_0	X_1	X_2	X_3
1	+	-	-	+
2	+	+	-	-
3	+	-	+	-
4	+	+	+	+

X

N	X_1X_2	X_1X_3	X_2X_3
1	+	+	-
2	-	-	+
3	-	+	-
4	+	-	+

X_1

	X_0	X_1	X_2	X_3
X_0	4	0	0	0
X_1	0	4	0	0
X_2	0	0	4	0
X_3	0	0	0	4

X'X

X_1X_2	X_1X_3	X_2X_3
0	0	0
0	0	4
0	4	0
4	0	0

$X'X_1$

$$(X'X)^{-1} = \tfrac{1}{4} I_4$$

$$A = (X_1'X_1)/4 = \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{matrix}$$

Plackett-Burman design $2^3//4$

Modèle postulé : $\eta = X\beta$

Modèle vrai : $\eta = X\beta + X_1\beta_1$

⇨ $B = (X'X)^{-1} X'Y$ $E[B] = ?$

$E[B] = \beta + A\beta_1$

$E[b_0] = \beta_0$

$E[b_1] = \beta_1 + \beta_{23}$

$E[b_2] = \beta_2 + \beta_{13}$

$E[b_3] = \beta_3 + \beta_{12}$

Plackett-Burman design $2^k//N$

Postulated model : $\eta = X\beta$

«true» model : $\eta = X\beta + X_1\beta_1$

⇒ $B = (X'X)^{-1} X'Y$ $E[B] = ?$

$E[B] = \beta + A\beta_1$

$E[b_0] = \beta_0$

$E[b_i] = \beta_i \pm a_{ij}\beta_{ij}$

Plackett-Burman designs

$a_{ij} = 0$

or

$a_{ij} = \pm 1$

$a_{ij} = 0$

or

$a_{ij} \neq \pm 1$

Geometric

$N \equiv 0 \pmod 8$ et $N = 4$

Non geometric

$N \equiv 0 \pmod 8$ et $N = 4$

Geometric Plackett-Burman designs

Postulated model : $\eta = X\beta$

«True» model : $\eta = X\beta + X_1\beta_1$

$B = (X'X)^{-1} X'Y$ $\quad E[B] = ?$

$E[B] = \beta + A\beta_1$

$2^{3}//4$

$E[b_0] = \beta_0$

$E[b_1] = \beta_1 + \beta_{23}$

$E[b_2] = \beta_2 + \beta_{13}$

$E[b_3] = \beta_3 + \beta_{12}$

$2^{7}//8$

$E[b_0] = \beta_0$

$E[b_1] = \beta_1 - \beta_{34} - \beta_{26} - \beta_{57}$

....

$E[b_7] = \beta_7 - \beta_{23} - \beta_{15} - \beta_{46}$

Non-geometric Plackett-Burman designs

Postulated model : $\eta = X\beta$

«true» model : $\eta = X\beta + X_1\beta_1$

$B = (X'X)^{-1} X'Y$ $\qquad E[B] = ?$

$\Rightarrow$ $E[B] = \beta + A\beta_1$

$2^{11}//12$

$E[b_0] = \beta_0$

$E[b_1] = \beta_1 \pm 0.33\, \beta_{m,j} \quad \ldots. \quad m, j \neq 1$

....

$E[b_{11}] = \beta_{11} \pm 0.33\, \beta_{m,j} \quad \ldots \quad m, j \neq 11$

$2^{82}//84$

$E[b_0] = \beta_0$

$E[b_j] = \beta_j \pm a_{m,j}\, \beta_{m,j} \quad \ldots. \quad m, j \neq 1$

$a_{m,j} = 0.04;\ 0.10;\ 0.20;\ 0.28$

Resolution

A design of resolution L is one in which no m-factor effect is confounded with any other effect containing less than L-m factors

Roman number

R_L

Plackett-Burman designs → R_{III}

Rechtschaffner designs → R_{V}

The probability that a factor is active is SMALL

k factors with 2 levels ($N \leq 100$)

Plackett-Burman designs R_{III}

$N \geq k + 1$

$N \equiv 0$ (modulo 4)

Plackett-Burman designs R_{IV}

$N \geq 2(k + 1)$

$N \equiv 0$ (modulo 4)

The probability that a factor is active is SMALL

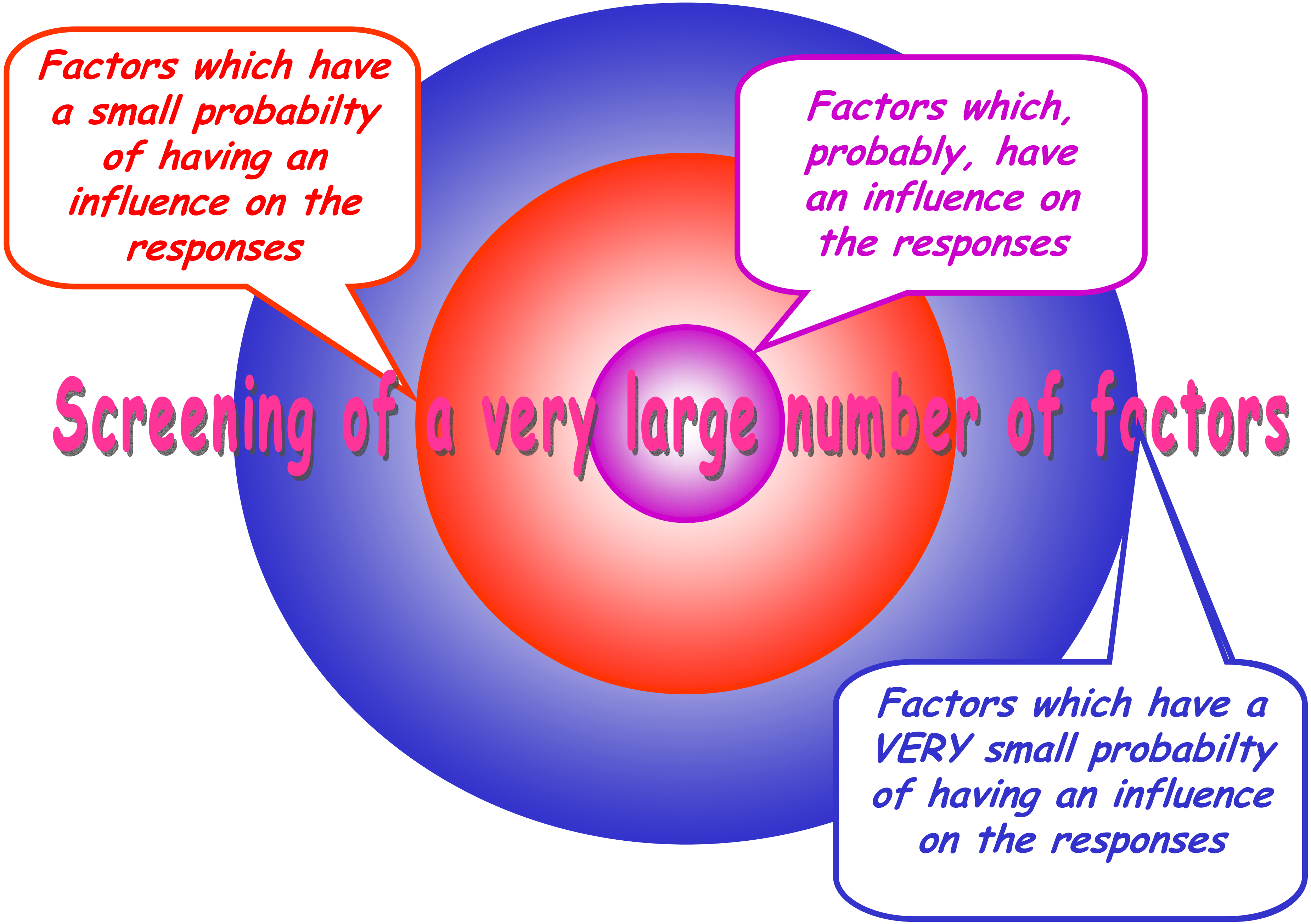
k factors with q_j levels ($N \leq 100$)

Symetrical or asymetrical designs

$$N \geq 1 + \Sigma(q_j - 1)$$

q^k // N

$2^a\ 3^b\ \ldots . 6^f$ // N


Factors which have a small probabilty of having an influence on the responses
Factors which, probably, have an influence on the responses
Screening of a very large number of factors
Factors which have a VERY small probabilty of having an influence on the responses

The probability that a factor is active is VERY SMALL

Supersaturated Designs $2^k//N$

$k >> N$

- some effects are predominant,
- the interactions are insignificant

SATTERTHWAITE (1959) :

Random procedure

For each factor :

- the levels of the factors are balanced,
- the distribution of the levels is obtained by random

SATTERTHWAITE

N	A_1	A_2	-	B_1	-	C_2	-	D_6	-	E_{12}	
1	1	4	-	3	-	4	-	3	-	2	
2	1	6	-	5	-	2	-	1	-	1	*A 5 six-level factors*
3	2	3	-	1	-	1	-	2	-	1	
4	4	5	-	4	-	4	-	2	-	2	*B 8 five-level factors*
5	6	6	-	2	-	3	-	3	-	2	
6	3	1	-	2	-	1	-	3	-	2	*C 9 four-level factors*
7	4	2	-	4	-	2	-	1	-	2	
8	5	1	-	5	-	3	-	2	-	1	*D 7 three-level factors*
9	6	3	-	4	-	1	-	1	-	1	
10	2	4	-	5	-	4	-	1	-	1	*E 11 two-level factors*
11	3	5	-	3	-	3	-	2	-	2	
12	5	2	-	1	-	2	-	3	-	1	$2^{11}\ 3^7\ 4^9\ 5^8\ 6^5$ // 12

Supersaturated designs

$k >> h$

- **k** : the number of potentially influential factors is very important
- **h** : the number of factors actually influential is probably very small
- The interaction effects are probably very small
- The number of experiments **N** is lower than the number of required informations **NI**

$$NI = 1 + \sum (q_j - 1)$$

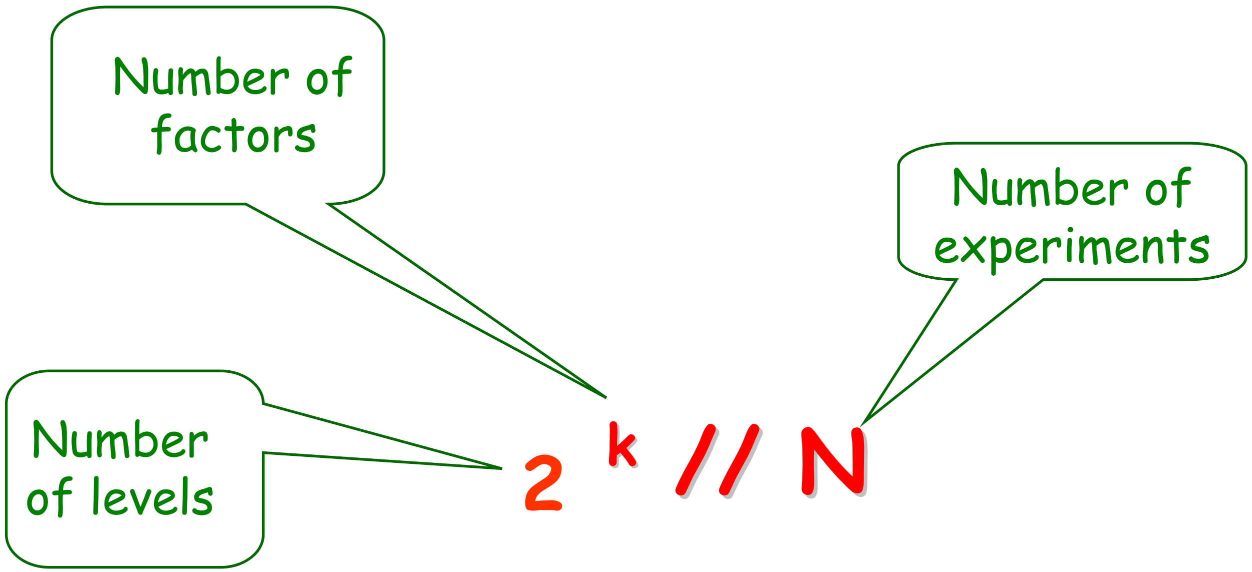
q_j : number of levels of factor j

Supersaturated designs

$k >> h$

- The most used supersaturated designs, *currently*, are those which all the factors have 2 levels

$$N < NI = 1 + 2k$$



BOOTH et COX (1962) :

7 supersaturated designs

$2^{16}//12$ $2^{20}//12$ $2^{24}//12$ $2^{24}//18$

$2^{30}//18$ $2^{36}//18$ $2^{30}//24$

CRITERIA

BOOTH and COX (1962) :

$$E(s^2) = \sum_{\substack{1 \le i,j \le k \\ i \ne j}} \langle d_i, d_j \rangle \Big/ \binom{k}{2}$$

d_i : column of the factor i, with the levels ± 1

$< d_i, d_j >$: cross product between columns d_i and d_j

$$s_{ij}^2 = \Sigma (d'_i, d_j)^2$$

$E(s^2) = \Sigma [s_{ij}^2] / (k! /[2! (k-2) !]$: $i < j$: 1, ..., k

$E(s^2)$: is a measure of the non-orthogonality.

N	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
1	+	-	+	+	+	-	-	-	+	-
2	-	+	+	+	-	-	-	+	-	+
3	+	+	-	-	-	+	-	+	+	-
4	+	-	-	-	+	-	+	+	-	+
5	-	-	-	+	-	+	+	-	+	+
6	-	+	+	-	+	+	+	-	-	-

							x_7
							-
							-
							-
							+
							+
							+
x_2	-	+	+	-	-	+	-2

$x'_2 x_7 = -2$

$X'X$

	X_0	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
X_0	6										
X_1	0	6									
X_2	0	-2	6								
X_3	0	-2	2	6							
X_4	0	-2	-2	2	6						
X_5	0	2	-2	2	-2	6					
X_6	0	-2	2	-2	-2	-2	6				
X_7	0	-2	-2	-2	-2	2	2	6			
X_8	0	2	2	-2	-2	-2	-2	-2	6		
X_9	0	2	-2	-2	2	-2	2	-2	-2	6	
X_{10}	0	-2	-2	-2	2	-2	-2	2	2	-2	6

Symmetry

CORRELATION MATRIX

	X_0	X_1	X_a	X_3	X_4	X_5	X_1	X_7	X_8	X_9	X_{10}
X_0	1										
X_1	0	1									
X_a	0	-a	1								
X_3	0	-a	a	1							
X_4	0	-a	-a	a	1						
X_5	0	a	-a	a	-a	1					
X_1	0	-a	a	-a	-a	-a	1				
X_7	0	-a	-a	-a	-a	a	a	1			
X_8	0	a	a	-a	-a	-a	-a	-a	1		
X_9	0	a	-a	-a	a	-a	a	-a	-a	1	
X_{10}	0	-a	-a	-a	a	-a	-a	a	a	-a	1

Symmetry

$a = 0.3333$

- **LIN** (1991, 1993, 1995, ...)
- **WU** (1993, ...)
- **NGUYEN** (1996)
- **YAMADA** (1997, ...)
- **CHENG** (1997)
- **DENG, LIN et WANG** (1999, ...)
- **LIU et ZHANG** (2000, ...)
- **CELA** (2000)

Construction methods

- Lin (1991, 1995) : algorithm
- Lin (1993) : from Plackett-Burman designs non geometrics
- Wu (1993) : from Plackett-Burman designs non geometrics
- Deng,Lin etWang (1994): from Plackett-Burman designs
- Liu et Zhang
- Nguyen (1996) : from BBID
- Yamada et Lin (1997) : from OA
- Cheng (1997) : algorithm
- Cela (1998) : genetic algorithm

LIN (1993):

- Half fractions of Plackett-Burman designs

(N-2) factors in N/2 runs

Plackett-Burman design $2^{11}//12$

N	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
1	+	+	-	+	+	+	-	-	-	+	-
2	+	-	+	+	+	-	-	-	+	-	+
3	-	+	+	+	-	-	-	+	-	+	+
4	+	+	+	-	-	-	+	-	+	+	-
5	+	+	-	-	-	+	-	+	+	-	+
6	+	-	-	-	+	-	+	+	-	+	+
7	-	-	-	+	-	+	+	-		+	+
8	-	-	+	-	+	+	-	+	+	+	-
9	-	+	-	+	+	-	+		+	-	-
10	+	-	+	+	-	+	+		-	-	-
11	-	+	+	-	+	+		-	-	-	+
12	-	-	-						-	-	-

Branching colums

Plackett-Burman design $2^{11}//12$

N	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
1	+	+	-	+	+	+	-	-	-	+	-
2	+	-	+	+	+	-	-	-	+	-	+
3	-	+	+	+	-	-	-	+	-	+	+
4	+	+	+	-	-	-	+	-	+	+	-
5	+	+	-	-	-	+	-	+	+	-	+
6	+	-	-	-	+	-	+	+	-	+	+
7	-	-	-	+	-	+	+	-	+	+	+
8	-	-	+	-	+	+	-	+	+	+	-
9	-	+	-	+	+	-	+	+	+	-	-
10	+	-	+	+	-	+	+	+	-	-	-
11	-	+	+	-	+	+	+	-	-	-	+
12	-	-	-	-	-	-	-	-	-	-	-

Branching colums

N	X_1	X_2	X_3	X_4						10	X_{11}
2	+	-	+	+	+	-	-	-		-	+
3	-	+	+	+	-	-	-	+	-		+
5	+	+	-	-	-	+	-	+	+	-	+
6	+	-	-	-	+	-	+	+	-	+	+
7	-	-	-	+	-	+	+	-	+	+	+
11	-	+	+	-	+	+	+	-	-	-	+

N	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
1	+	+	-	+	+	+	-	-	-	+	-
4	+	+	+	-	-	-	+	-	+	+	-
8	-	-	+	-	+	+	-	+	+	+	-
9	-	+	-	+	+	-	+	+	+	-	-
10	+	-	+	+	-	+	+	+	-	-	-
12	-	-	-	-	-	-	-	-	-		-

Branching colums

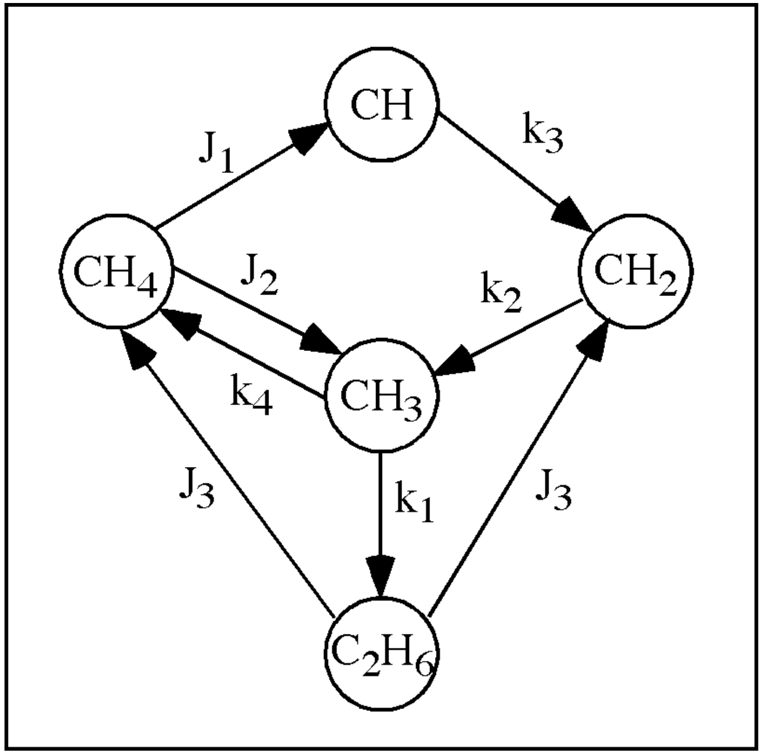
Criteria of quality

- $E(s^2)$ ⇒ min $E(s^2)$
- $|r\ max|$ (r : non-diagonal correlation matrix elements) ⇒ min $|r\ max|$
- $\max |s_{ij}|$ (s_{ij} : non-diagonal information matrix elements) ⇒ $\min \max |s_{ij}|$
- Frequency $\max |s_{ij}|$
- ....

ANALYSIS

- Ridge regression
- Step-Wise regression
- All-subset regression
- genetic algorithm
- Approach bayesian

Modélisation de la photochimie du méthane dans l'atmosphère de Titan et des planètes géantes.



$$CH_4 + h\nu \rightarrow CH + H_2 + H \quad q_1 = 0.51 \quad J_1$$
$$CH_4 + h\nu \rightarrow CH_3 + H \quad q_2 = 0.49 \quad J_2$$
$$CH_3 + CH_3 + M \rightarrow C_2H_6 + M \quad k_1$$
$$C_2H_6 + h\nu \rightarrow CH_4 + CH_2 \quad J_3$$
$$CH_2 + H_2 \rightarrow CH_3 + H \quad k_2$$
$$CH + H_2 \rightarrow CH_2 + H \quad k_3$$
$$CH_3 + H + M \rightarrow CH_4 + M \quad k_4$$

M. Dobrijevic, M. Claes-Bruno, M. Sergent, R. Phan-Tan-Luu

$$\frac{d[\mathrm{CH_4}]}{\mathrm{dt}} = J_3[\mathrm{C_2H_6}] - (J_1 + J_2)[\mathrm{CH_4}] + k_4[\mathrm{H}][\mathrm{CH_3}][\mathrm{M}]$$

$$\frac{d[\mathrm{C_2H_6}]}{\mathrm{dt}} = k_1[\mathrm{CH_3}]^2[\mathrm{M}] - J_3[\mathrm{C_2H_6}]$$

$$\frac{d[\mathrm{CH}]}{\mathrm{dt}} = J_1[\mathrm{CH_4}] - k_3[\mathrm{H_2}][\mathrm{CH}]$$

$$\frac{d[\mathrm{CH_3}]}{\mathrm{dt}} = J_2[\mathrm{CH_4}] + k_2[\mathrm{H_2}][\mathrm{CH_2}] - k_1[\mathrm{CH_3}]^2[\mathrm{M}] - k_4[\mathrm{H}][\mathrm{CH_3}][\mathrm{M}]$$

$$\frac{d[\mathrm{CH_2}]}{\mathrm{dt}} = k_3[\mathrm{H_2}][\mathrm{CH}] + J_3[\mathrm{C_2H_6}] - k_2[\mathrm{H_2}][\mathrm{CH_2}]$$

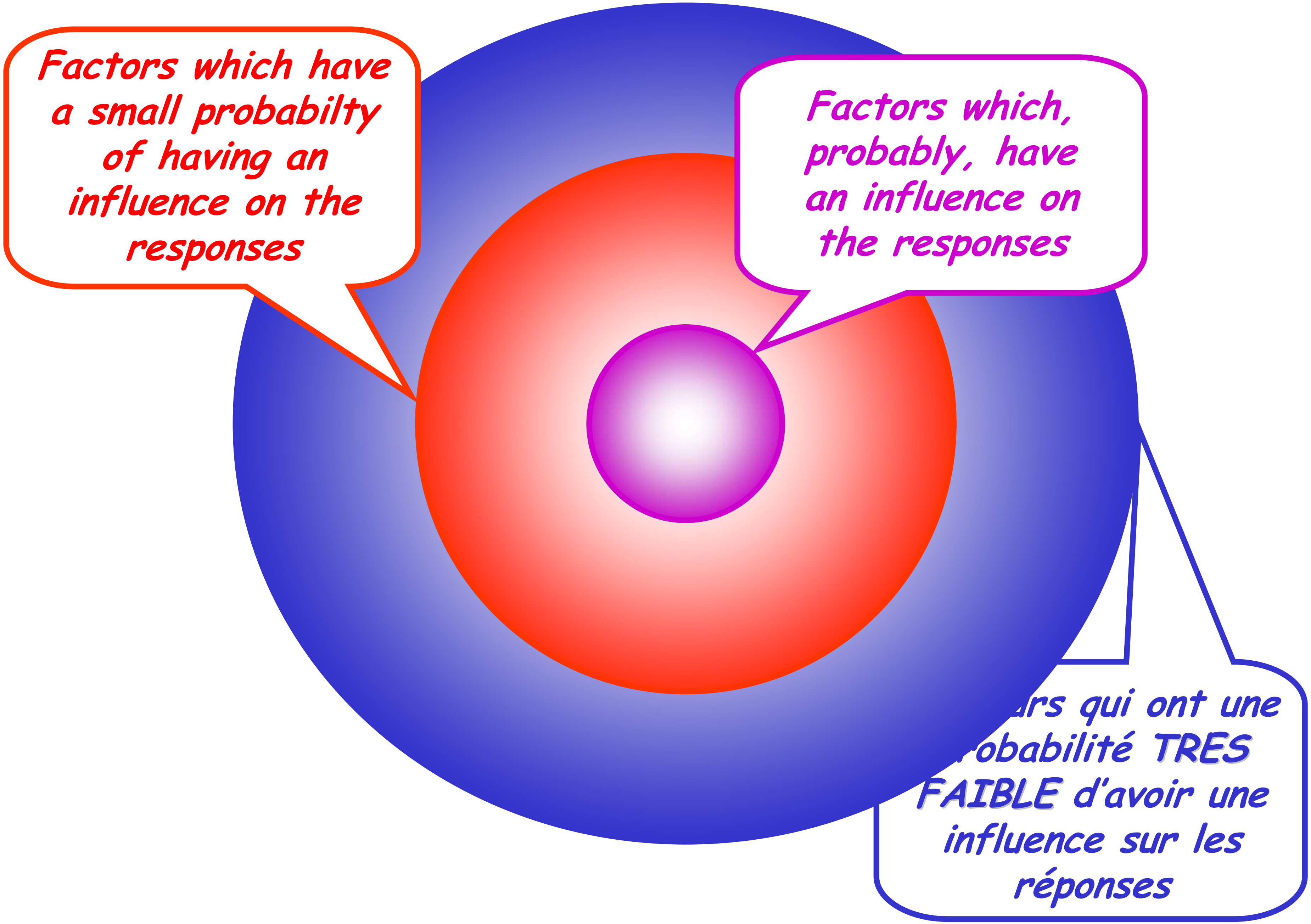
$$\frac{d[\mathrm{H}]}{\mathrm{dt}} = (J_1 + J_2)[\mathrm{CH_4}] + k_3[\mathrm{H_2}][\mathrm{CH}] + k_2[\mathrm{H_2}][\mathrm{CH_2}] - k_4[\mathrm{H}][\mathrm{CH_3}][\mathrm{M}]$$

$$\frac{d[\mathrm{H_2}]}{\mathrm{dt}} = J_1[\mathrm{CH_4}] - k_3[\mathrm{H_2}][\mathrm{CH}] - k_2[\mathrm{H_2}][\mathrm{CH_2}]$$

82 Factors with 2 levels **15 reponses**

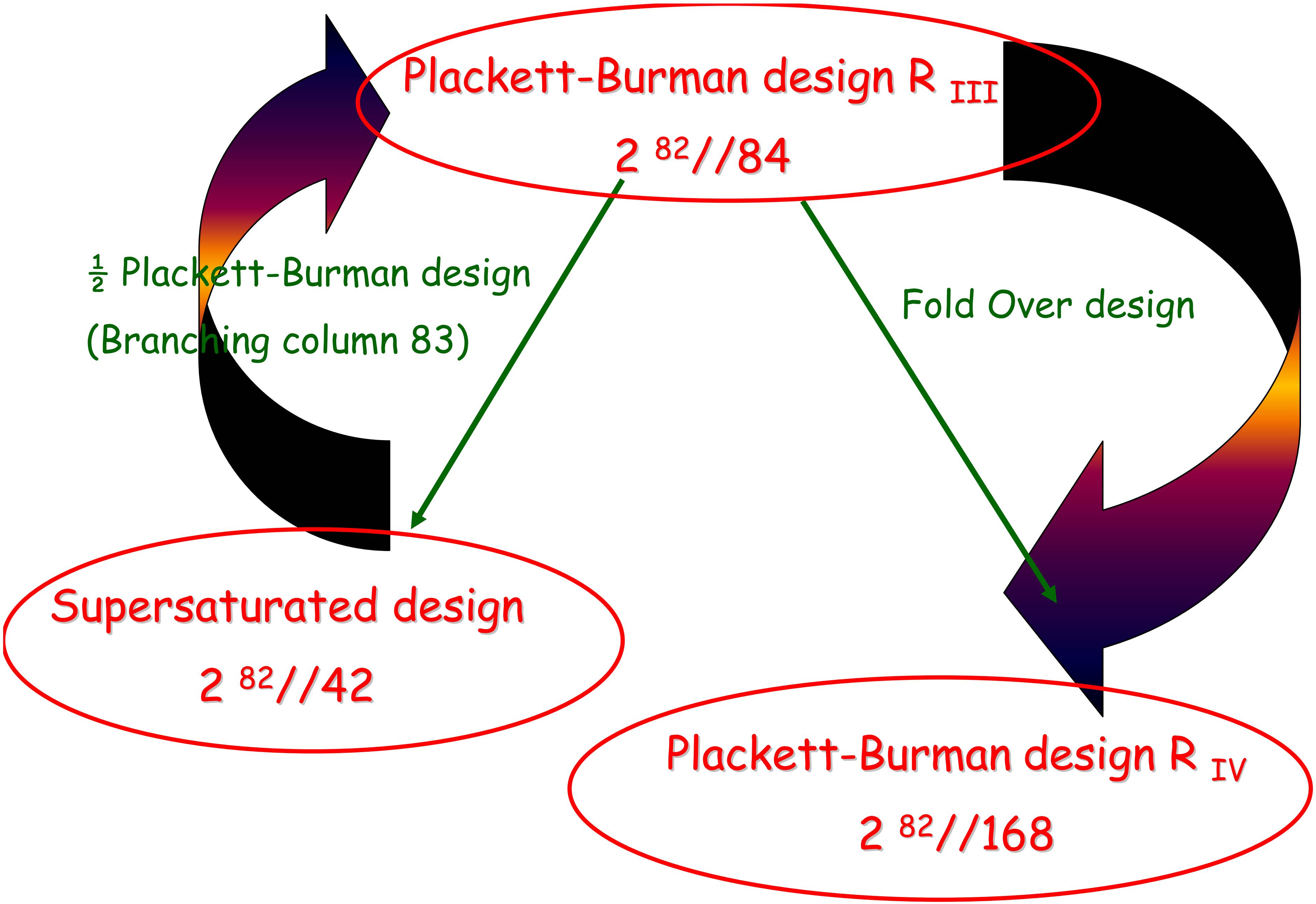
(constantes de réaction) **(Concentration de 15 composés)**

M. Dobrijevic, J.L. Ollivier, F. Billebaud, J. Brillet, J.P. Parisot.
Effect of chemical kinetics uncertainties on photochemical
modeling results: application to Saturn's atmosphere.
Astronomy and Astrophysics. Vol. 398, 335-344, 2003.


Factors which have a small probabilty of having an influence on the responses
Factors which, probably, have an influence on the responses
urs qui ont une
obabilité TRES FAIBLE d'avoir une influence sur les réponses

The probability that a factor is active is

VERY SMALL	~~Sequential bifurcation~~ ~~Group Screening~~ Supersaturated designs
SMALL (no interaction effects)	Plackett-Burman designs R_{III}
SMALL (interaction effects possible)	Plackett-Burman designs R_{IV}



Plackett-Burman design R III
2 82//84
½ Plackett-Burman design
(Branching column 83)
Fold Over design
Supersaturated design
2 82//42
Plackett-Burman design R IV
2 82//168

Quantitative study of factors

↓

We take into account the possibility of **interaction** effects between the factors

↓

The effect of a factor can be different according to the value of another factor

Rechtschaffner designs

k factors with 2 levels
All main effects (β_i) and all interaction effects (β_{ij})
$N = 1 + k(k-1)/2$

Rechtschaffner designs ⇒ R_V

	X_1	X_2	X_3	X_4	….	….	X_{k-1}	X_k
1	-1	-1	-1	-1	….	….	-1	-1
k	-1	+1	+1	+1	….	….	+1	+1
k*(k-1)/2	+1 +1 …. -1	+1 -1 …. -1	-1 +1 …. -1	-1 -1 …. -1	…. …. …. ….	…. …. …. ….	-1 -1 …. +1	-1 -1 …. +1

To elaborate the experimental strategy

to choose an appropriate Experimental design in accordance with

THE TARGETS

To know, anywhere in the experimental domain, the value of the respones

Optimization

Design of experiments

Optimization

To look for the optimum of one or more responses in the experimental domain of interest

↓

To know, in every point of the domain, the value of the response(s)

Continuous domain

Small domain

Metamodel possible

- Doehlert designs
- Hybrid designs
- Composite designs,
- …

RSM designs

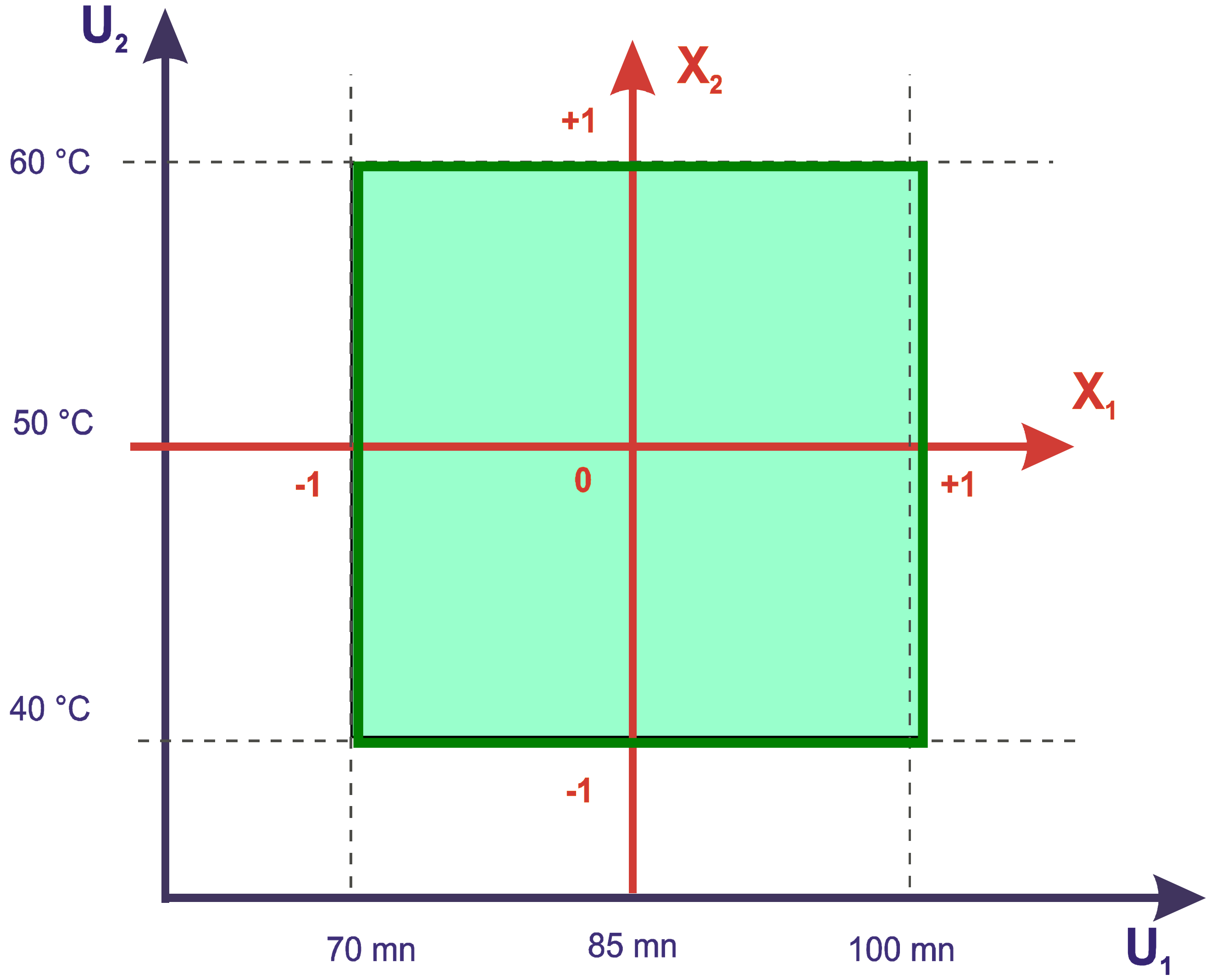
Large domain

Metamodel not possible

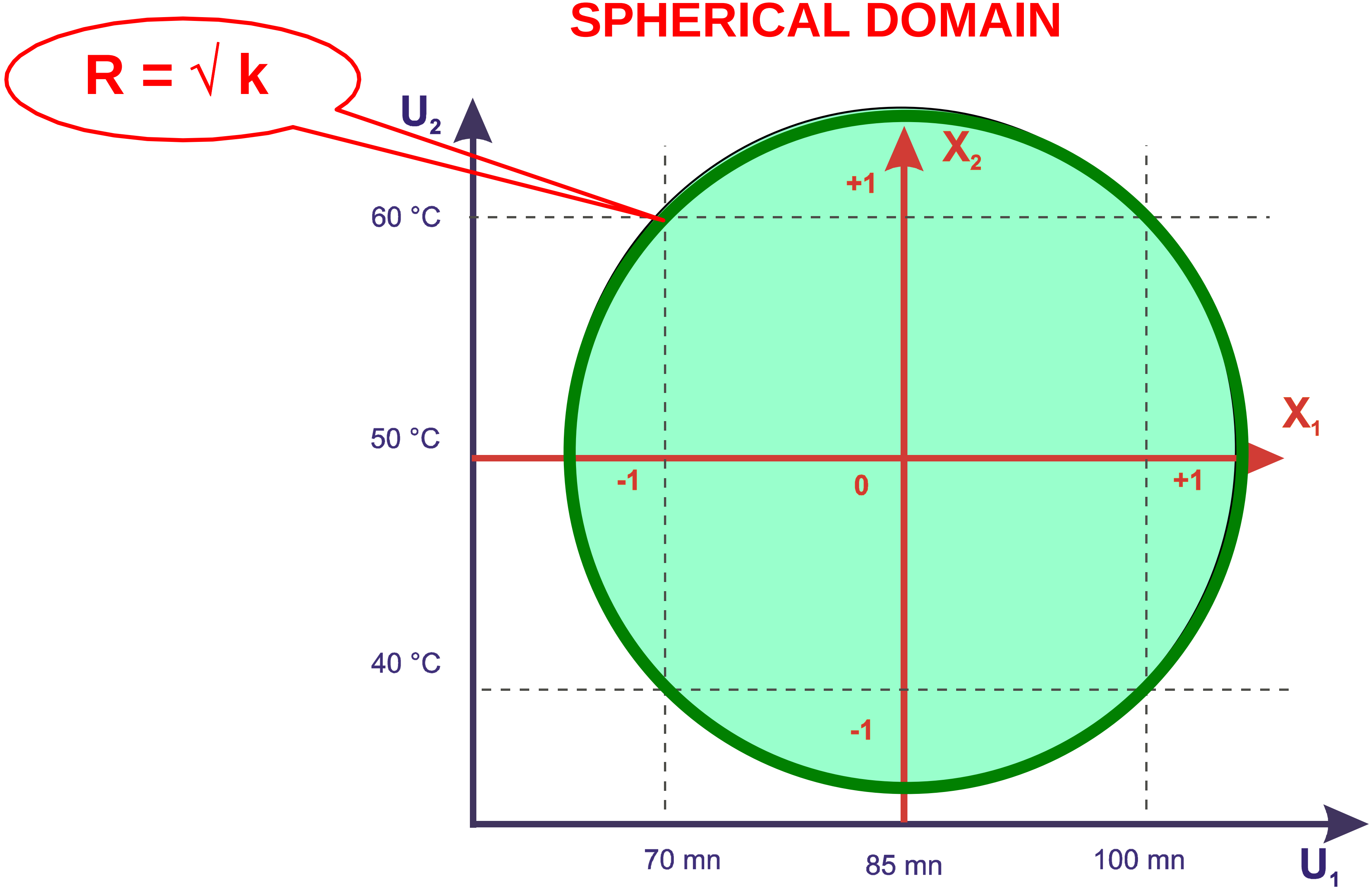
- Latin Hypercubes
- Uniform designs
- Hammersley, Halton,..
- …

Space-filling designs

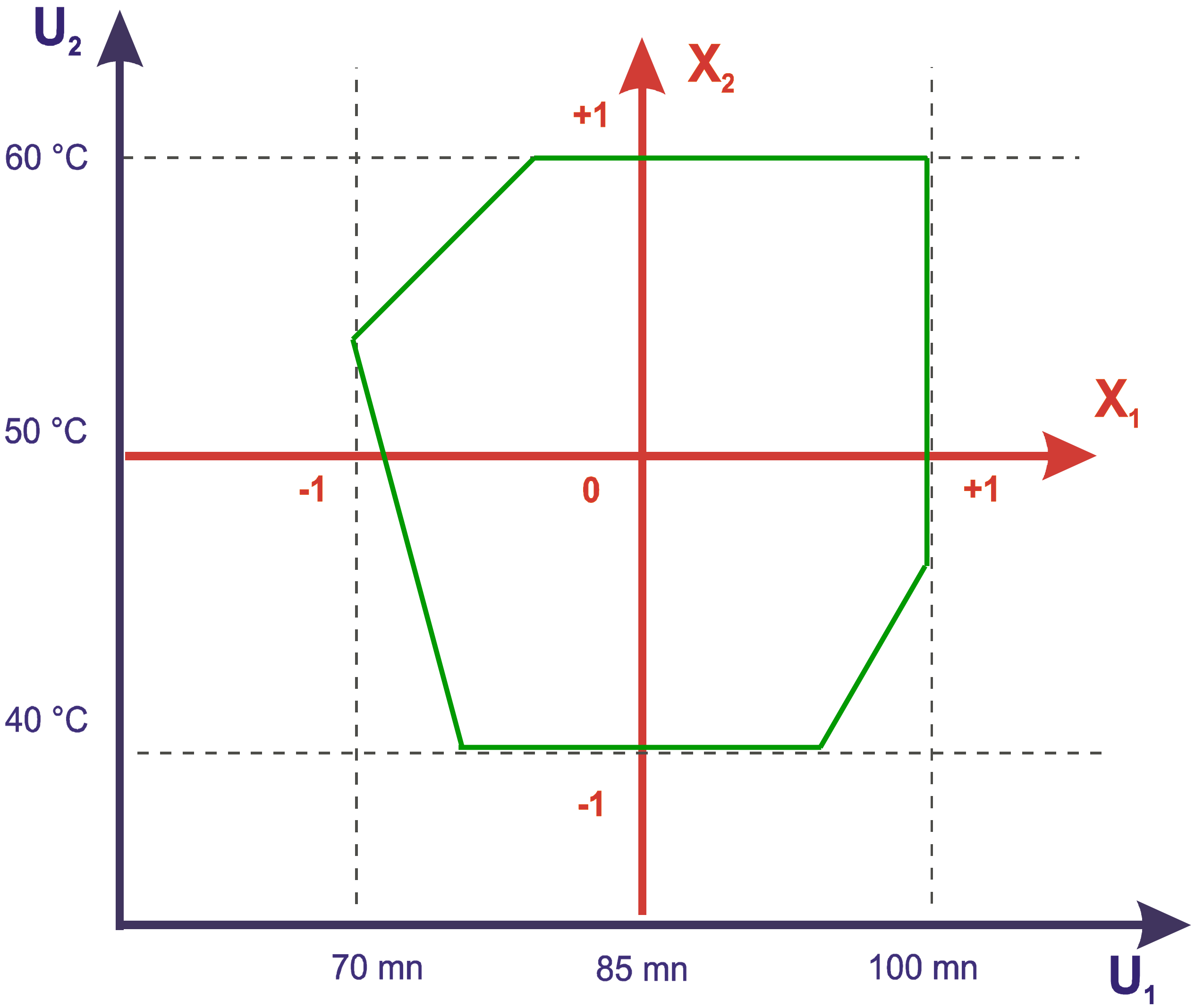
CUBICAL DOMAIN



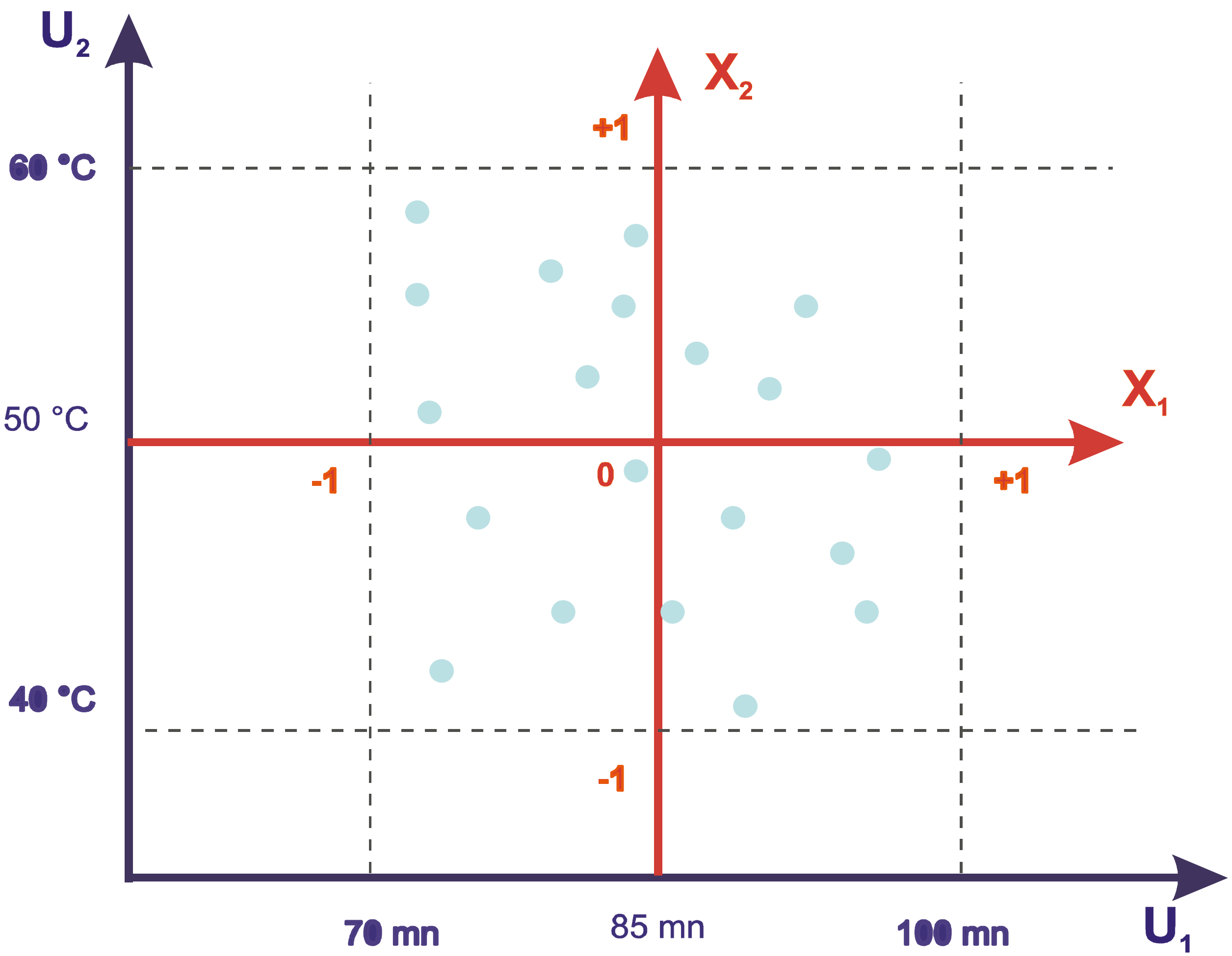
SPHERICAL DOMAIN



CONSTRAINT DOMAIN



CONSTRAINT DOMAIN



What do we want ?

To know, anywhere in the domain of interest, the value of the responses

To find, if it exists, the domain where all the experimental responses respect the constraints imposed by the specifications

zone of acceptable compromise

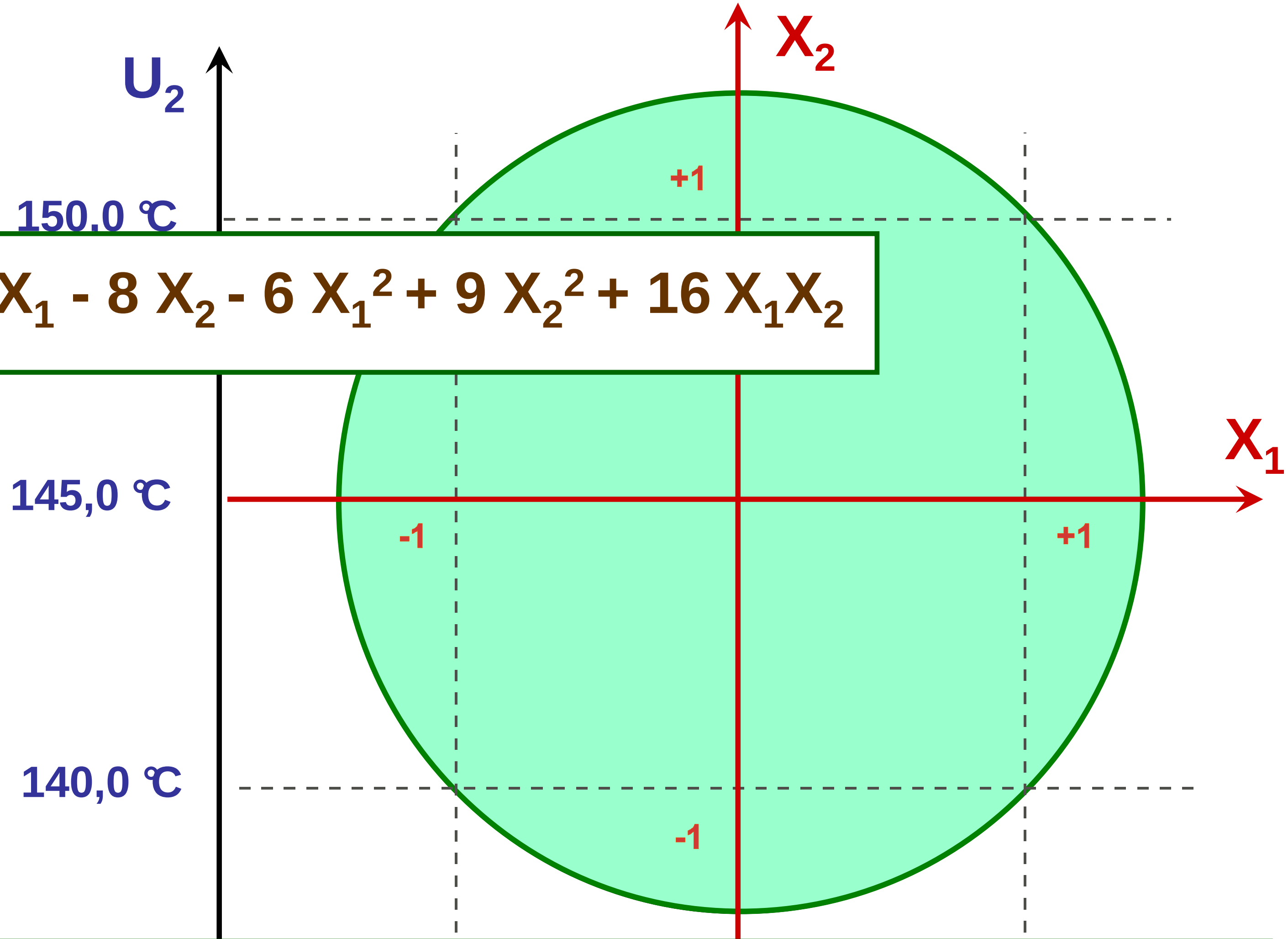
We must do experiments

Which ones ?

Those, bringing the desired information !

Infinity of experiments
(or simulations) !

$$Y = 55 + 12\,X_1 - 8\,X_2 - 6\,X_1^2 + 9\,X_2^2 + 16\,X_1X_2$$



$$\eta = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_{11}X_1^2 + \beta_{22}X_2^2 + \beta_{12}\,X_1X_2$$

We must do experiments

Which ones ?

Those, bringing the desired information !

Mathematical model

Infinity of experiments
(or simulations) !

GOAL

To know, anywhere in the domain of interest, the value of the responses

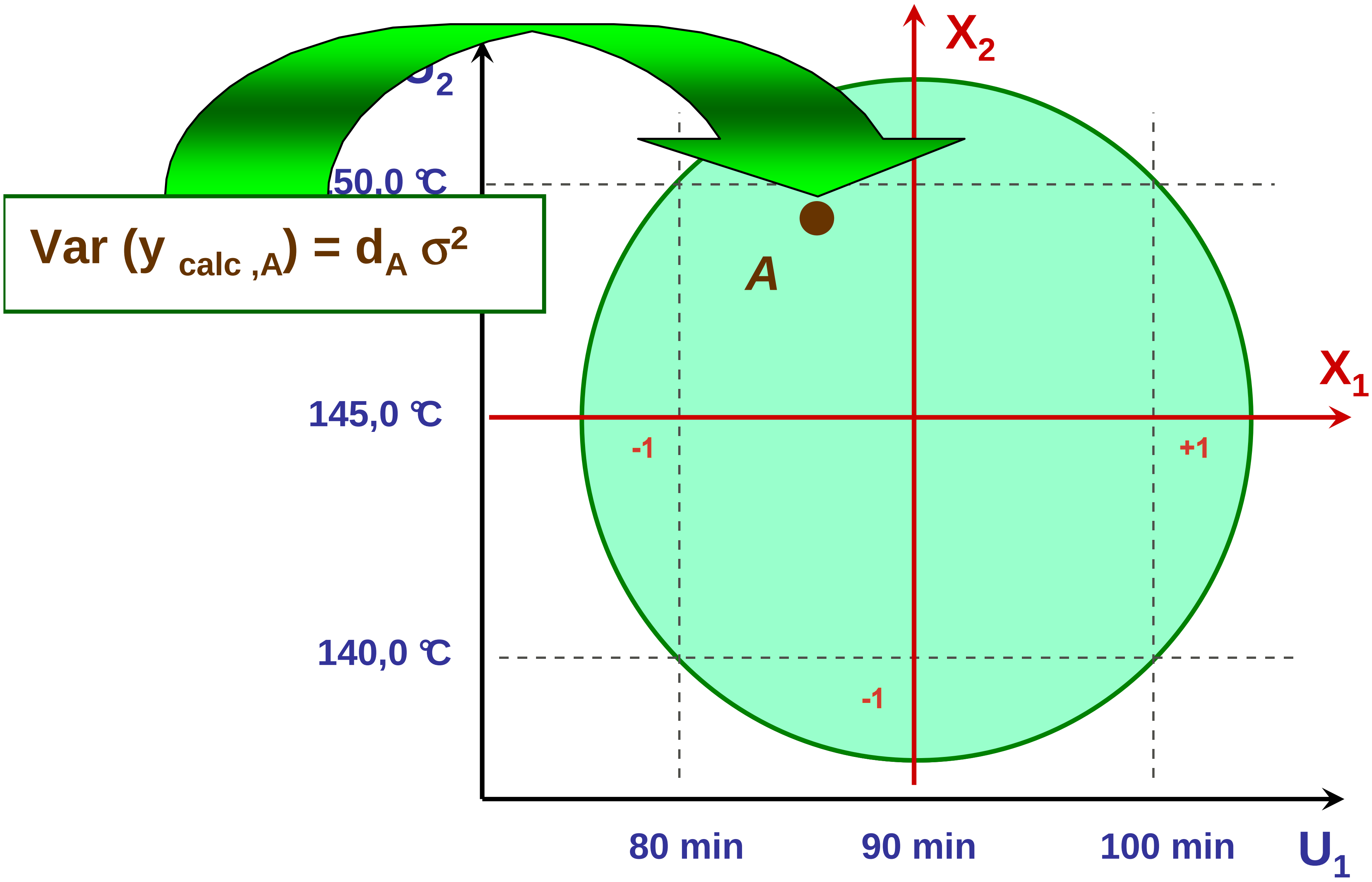
To postulate a model

Which are the qualities of the model ?

The model must well represent the phenomenon in the domain of interest

It must predict the response, anywhere in the domain, with a good predictive quality

It must predict the response, anywhere in the domain, with a good predictive quality



$$\eta_A = \beta_0 + \beta_1 x_{A1} + \beta_2 x_{A2} + \beta_{11} x_{A1}^2 + \beta_{22} x_{A2}^2 + \beta_{12}\, x_{A1} x_{A2}$$

x'_A : $\{1,\ x_{A1},\ x_{A1},\ x_{A1}^2,\ x_{A2}^2,\ x_{A1}x_{A2}\}$

$$y_{calc,A} = x'_A\ B$$

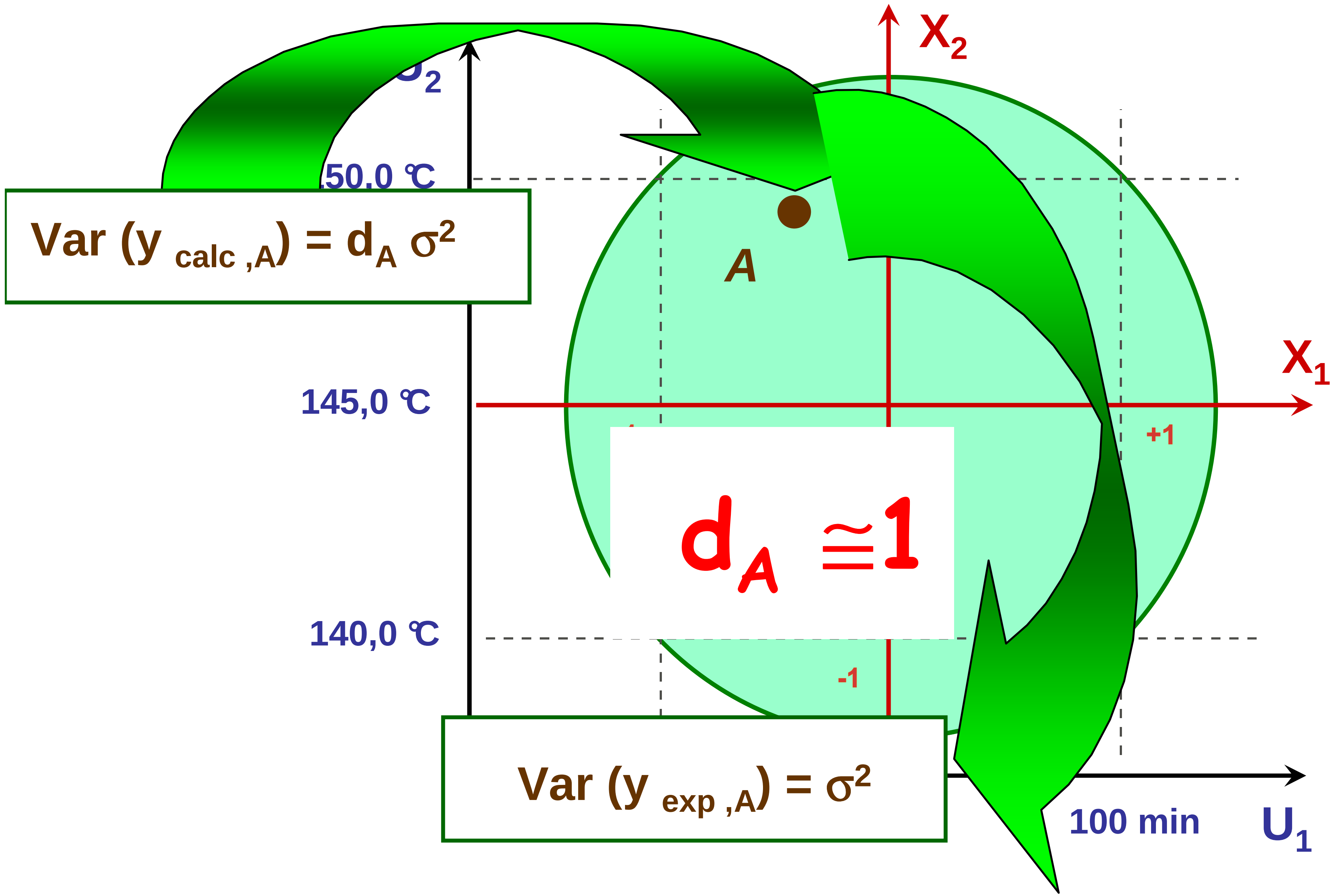
$$\mathrm{var}\ (y_{calc,A}) = \mathrm{var}\ (x'_A\ B)$$

$$= x'_A \mathrm{var}\ (B)\ x_A = x'_A\ (X'X)^{-1}\ x_A\, \sigma^2$$

$$= d_A\, \sigma^2$$

d_A : variance function

It must predict the response, anywhere in the domain, with a good predictive quality



What do we want ?

If the model is proved, it should allow to predict, in any point of this experimental domain, the value of the experimental response with the same quality we would have had if we made the experiment in the same point.

$$d_{Max} \cong 1$$

GOAL

To know, anywhere in the domain of interest, the value of the responses

To postulate a model

Construction of the experimental strategy

no

Validation of the model

yes

GOAL

To know, anywhere in the domain of interest, the value of the responses

To postulate a model

Full quadratic model

$$\eta = \beta_0 + \sum_i \beta_i x_i + \sum_{ii} \beta_{ii} x_i^2 + \sum_{i<j} \beta_{ij} x_i x_j$$

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

All the models are false,

but some of them are more useful than others.

GOAL

To know, anywhere in the domain of interest, the value of the responses

To postulate a model

- RSM designs

Construction of the experimental strategy

Validation design

Validation of the model

Response Surface Methodology

RSM DESIGNS

- ***Full quadratic model***

Spherical domain :

- composite (k < 12)
- Doehlert uniform shell
- hybrid designs (k < 7)
- Box- Benhken designs (k < 8)

RSM DESIGNS

- ***Full quadratic model***

  Cubical domain :

  - composite designs (k < 11)
  - Hoke designs (k < 6)

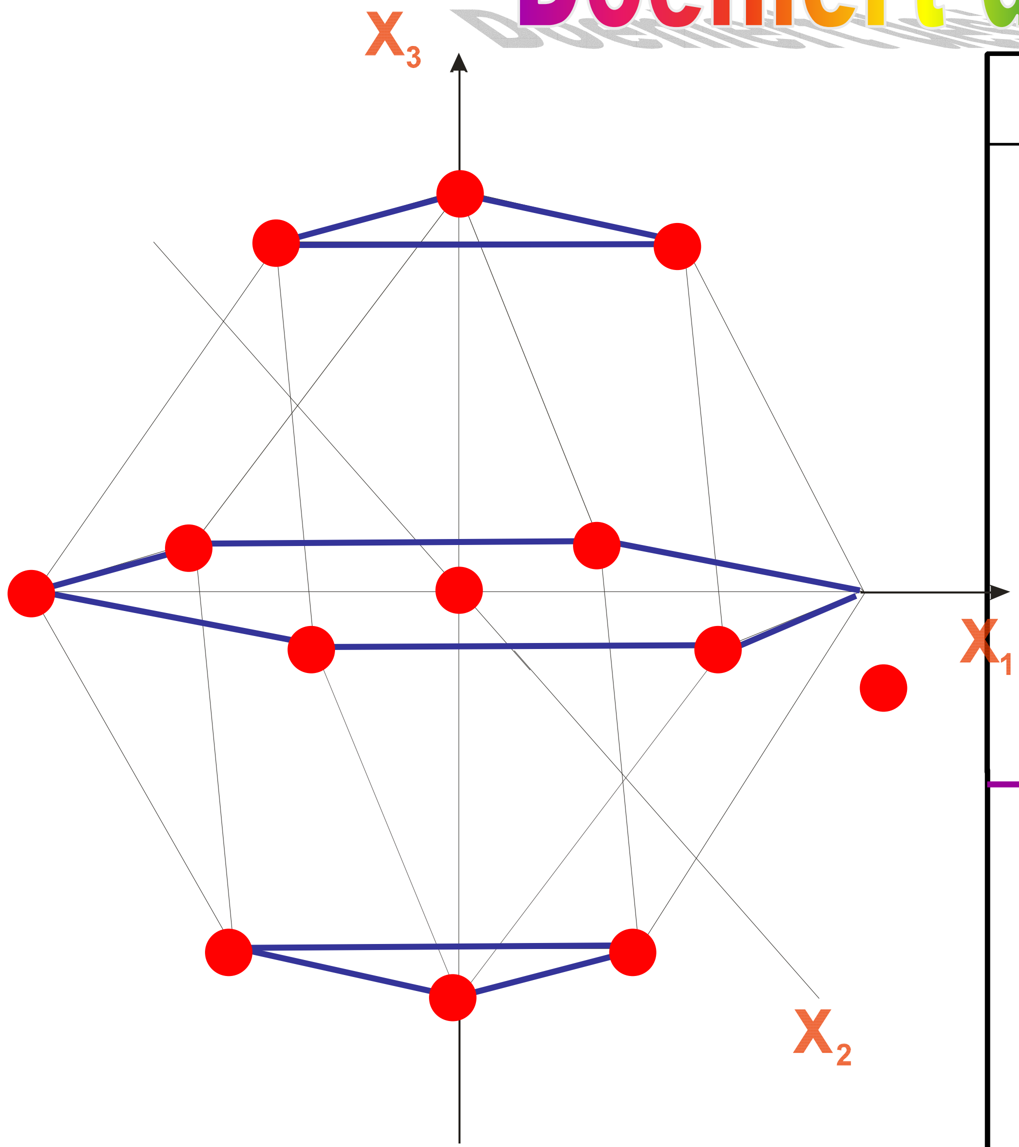
RSM DESIGNS

- ***Full quadratic model***

Constraint domain :

- generation of designs

Doehlert designs



N	X_1	X_2	X_3
1	0,000	0,000	0,000
2	1,000	0,000	0,000
3	0,500	0,866	0,000
4	- 1,000	0,000	0,000
5	- 0,500	- 0.866	0,000
6	0,500	- 0.866	0,000
7	- 0,500	0.866	0,000
8	0,500	0,289	0,816
9	- 0,500	- 0,289	- 0,816
10	0,500	- 0,289	- 0,816
11	0,000	0,577	- 0,816
12	- 0,500	0,289	0,816
13	0,000	- 0,577	0,816

Doehlert designs

k	p	N	ΔN	k	p	N	ΔN
2	6	7	6	13	105	183	28
3	10	13	8	14	120	211	30
4	15	21	10	15	136	241	32
5	21	31	12	16	153	273	34
6	28	43	14	17	171	307	36
7	36	57	16	18	190	343	38
8	45	73	18	19	210	381	40
9	55	91	20	20	231	421	42
10	66	111	22	21	253	463	44
11	78	133	24	22	276	507	

$$N_k = k^2 + k + 1 = N_{k-1} + 2k$$

Metamodel

↓

Experimental design

↓

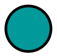
Model validation !

↓

Test points
uniformly scattered

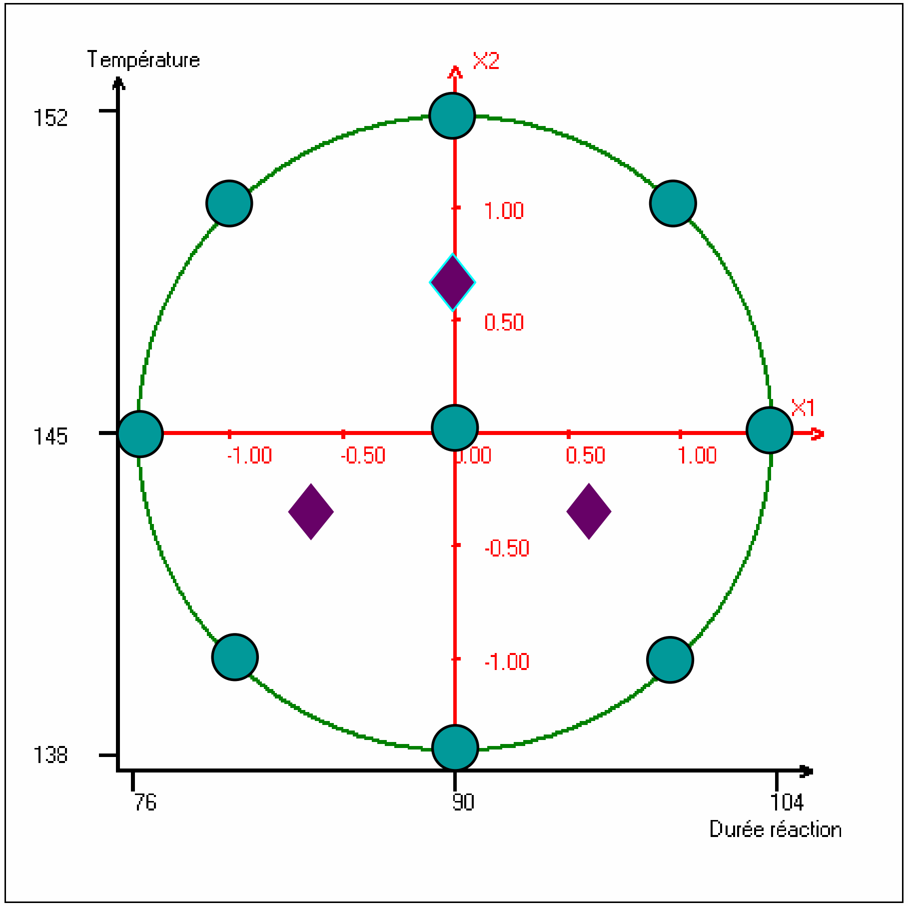
Space-filling designs

Validation design



RSM design

Test points



D.O.E

= "Hammer "

When one types oneself on the fingers

It is never the fault of the hammer !!!

Laboratoire de Méthodologie de la Recherche Expérimentale

UMR 6171 - Université Paul Cézanne - MARSEILLE

Roger Phan-Tan-Luu
Professeur

Michelle Sergent
Professeur

Magalie Claeys-Bruno
Maître de Conférences

e-mail : roger.phan-tan-luu@univ-cezanne.fr