الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur
et de la Recherche Scientifique
Université Akli Mohand Oulhadj - Bouira -
ⵜⴰⵙⴷⴰⵡⵉⵜ ⴰⴽⵍⵉ ⵎⵓⵃⴻⵏⴷ ⵓⵍⵃⴰⵊ - ⵜⵓⴱⵉⵔⴻⵜ -

جامعة البويرة

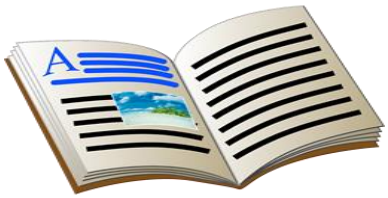
وزارة التعليم العالي والبحث العلمي
جامعة أكلي محند أولحاج
- البويرة -

Faculté des Sciences et des Sciences Appliquées

كلية العلوم والعلوم التطبيقية

Département de Génie des Procédés

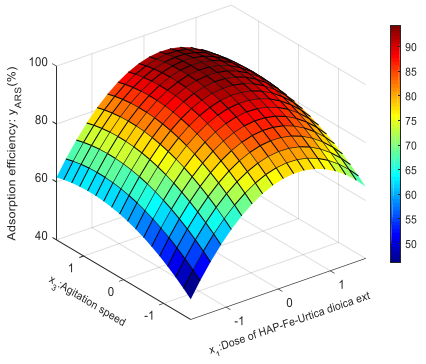
Polycopié de cours



En : Génie des Procédés

Spécialité : Génie Chimique et Génie de l'Environnement

Niveau : Master II



Principles and Practices of Experimental Designs

Par : Dr ZIANI Salima

Année : 2024/2025

Table de matière

Chapter III : *Statistical Analysis : Significance testing and model validation*

Chapter IV: *Fractional Factoriel Design*

Chapter V: Response surfaces methodology

Chapter VI: Mixture Designs

Preface

Although experimentation has existed as long as science itself, the concept of experimental design, specifically developed to optimize the process of data collection, only dates back to the 20th century. This modern methodology, refined over recent decades, has established itself as a reliable, universally applicable, and indispensable tool, allowing researchers to identify optimal conditions in data analysis with precision. The experimental design approach is, in fact, a structured methodology that facilitates planned experimental research. It assists researchers in organizing their investigations more systematically, validating their hypotheses with greater accuracy, achieving a clearer understanding of complex phenomena, and addressing specific research challenges effectively.

Experimental design is a systematic, economical, and rational method for organizing experiments involving multiple parameters. It ensures reliability and quality of results while maximizing information from a minimal number of trials. This methodology has gained success due to competitive demands and evolving experimental practices. However, is important to note that experimental design is not primarily designed for fundamental research as it cannot fully explain the physico-chemical phenomena under study.

This course document is designed for License's and Master's students, with a focus on Master's II students in Process Engineering (Chemical, Pharmaceutical and Environmental Engineering), Food Engineering, and Water Engineering and Management. It provides an in-depth analysis of experimental design, methodologies, statistical tools, and practical applications. The document also serves as a valuable resource for doctoral students, offering a solid theoretical framework and concrete examples for designing and analyzing experiments in their own research. However, it is recommended that students master mathematical and computational tools in order to effectively manage the statistical planning of experiments.

This document untitled « Principles and Practices of Experimental Designs », divided into six chapters, covers fundamentals and main experimental designs in various disciplines. It aims to provide experimenters with a procedure to optimize information extraction

while minimizing trials, addressing challenges in the current economic context. This mastery is crucial for those undertaking laboratory tests or industrial processes.

Chapter I
Concept and Approach of Experimental Design

Chapter I : Concept and Approach of Experimental Design

I. 1. Introduction

Experimentation is a fundamental tool across many scientific and engineering disciplines, providing researchers with the means to explore, deepen understanding, and address specific questions about various phenomena. In studying and monitoring the evolution of a process whether in Chemical Engineering, Environmental engineering, Polymer Engineering, Civil Engineering, Mechanical Engineering, Electrical Engineering, Physics, or Biological Engineering the experimenter can adopt a rigorous methodological approach that requires conducting a large number of trials. For example, when studying the impact of operational parameters on the yield of reactions in chemical engineering, or analysing the responses of materials subjected to various environmental conditions in civil engineering, the experimenter may have to carry out a number of tests based on two contradictory trends :

1. on the one hand, it is necessary to include all the factors that influence the process and to study them at different levels in order to obtain a representative model;
2. on the other hand, it is necessary to minimise the number of tests to reduce the duration and cost of the experiment.

Under these conditions, the methodological approach necessarily involves a comprehensive experimental strategy, encompassing both the planning and execution stages of the experiments.

I.2. Methodologies used for the optimal experimental strategy

The optimal strategy is one that, with a minimum of trials, provides an experimental model indicating the path to follow to reach the optimum. There are two primary strategies for achieving optimal experimental modelling :

I.2.1. Classical Gauss-Seidel method

The classic strategy, based on the ‘one factor at a time’ principle, involves holding all but one of the experimental factors constant, and the response (or objective function) is measured as a function of the variation in this single factor. This method makes it possible to observe the individual effect of each parameter on the result, but it has severe

limitations such as involving multiple factors with multiple experimental points, sometimes exceeding reasonable limits, and using the same algorithm for each parameter separately.

I.2.2. Modern Box-Wilson method : Experimental Design

The modern method, based on the principle of 'all factors at all times', involves varying the levels of all the factors simultaneously, in a planned and considered way. This strategy developped by **Ronald A. Fisher** (in the early 20th century) enhances understanding, efficiency, and precision in experiment planning, especially in areas where factor interactions significantly impact system dynamics. Its advantages include minimizing trial number, studying a large number of factors, identifying possible interactions, optimizing operating parameters, improving results accuracy compared to conventional models, and modeling results through mathematical model design.

I.3. Experimental Design approach

Experimental design is a systematic, efficient method and optimised strategy aimed at predicting, with maximum accuracy, a response based on a minimum number of experimental trials in line with the chosen theoretical model. This methodology enables scientists and engineers to study the relationship between multiple input variables ($\boldsymbol{x_i}$) and key output variables (or responses : $\boldsymbol{y_i}$), by modelling this relationship using a suitable mathematical.

$$y \quad = f(x_j)$$

With :

y : measured response or output variable ;

x_i: factors or system input variables.

Constructing an experimental design involves selecting a sufficient number **N** of specific combinations from the experimental space to reliably estimate the **M** unknown parameters of the model. This approach minimizes and standardizes uncertainty while closely adhering to the technical and economic constraints of the study.

I.4. Experimental design principle

The application of experimental design to simple and complex systems provides insights into their behaviour by treating them as a **black box** (Figure I.1). This approach involves

systematically varying input factors, observing the resulting changes in output responses, and inferring the relationships between inputs and outputs. Using this method, it is possible to infer how the system works without needing complete knowledge of the internal mechanisms driving the phenomenon. This allows effective analysis even when the underlying processes are complex or only partially understood.

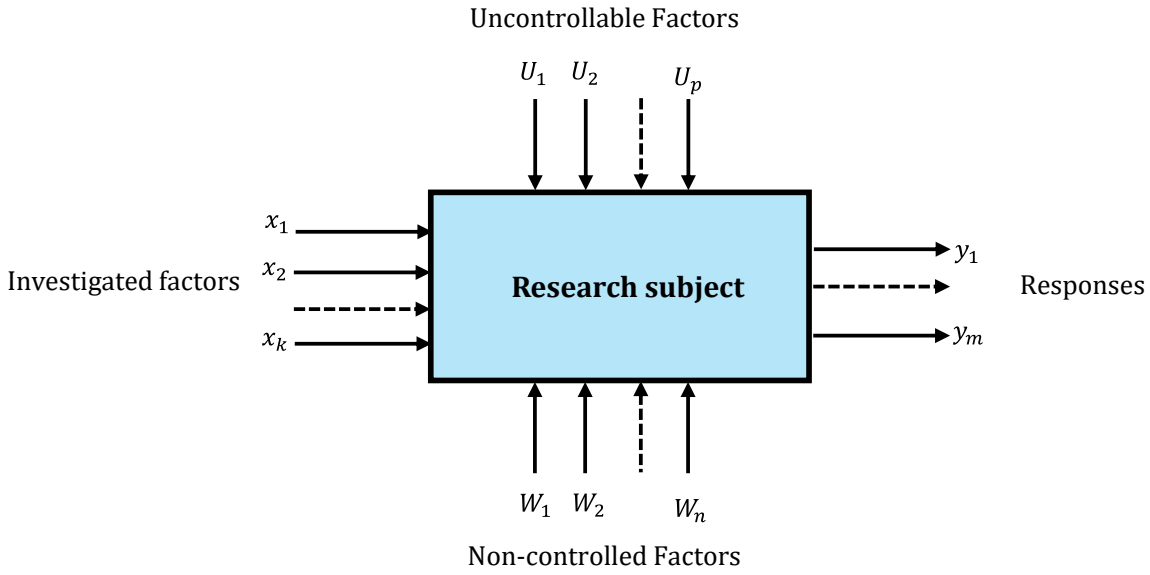


Figure I.1. Context of the black box model.

Input factors are controllable, uncontrollable and disturbance variables. Controllable factors are variables that can be directed, while uncontrollable factors are measured and controlled during the experiment but they cannot be changed at our wish. Non-controlled factors are disturbance variables, which are immeasurable and their values are randomly changed in time (e.g. deviations around the set points of the controllable factors, sampling and measurement error, system variability, etc.).

I.5. Chronological stages of experimental design

Experimental design emphasizes precise methodology programming through a sequence of stages to complete a study. The process involves formalizing the problem (defining objectives, factors, responses, experimental field), planning the experimental plan (construction of the experimental matrix), conducting experiments, analyzing the results, and optimizing the response by finding the optimal conditions for the study, ensuring that the objectives, factors, responses, and experimental field are clearly defined.

I.6. Terminology

Experimental design, like any other scientific discipline, has its own methodology, subject of research and terminology. The following section, highlighting the most crucial terminologies in this field.

I.6.1. Factors or real variables

Variables that can be adjusted by the experimenter and are intended to influence the response. These variables noted $\boldsymbol{z_j}$ can be of different types :

- **Continuous factors :** they can take on any real numerical value in the range studied, for example: process operating parameters such as temperature (20, 30, 40°C), pressure (0.5, 1, 1.5 atm), concentration (10, 50, 100 g/L), pH (2, 4, 6), flow rate (2, 3, 4L/min), etc. ;
- **Discrete factors :** they can only take specific values, such as a name, a letter, or a particular attribute. These discrete variables include **boolean factors**, wich can only take two values (e.g. high or low, open or closed, white or black, etc.) and **orderable factors**, wich describe a logical order (e.g. large, medium, small, or first, second, third and fourth, etc.).

The value assigned to a factor during an experimental trial is called a **level**. The study of the influence of a factor usually involves limiting its variations within two limits (Figure I.2).

- the low limite, which represents the low level : $z_{j(low)}$
- the upper limit, which represents the high level : $z_{j(upp)}$.

However, the factor's domain refers to the set of values that a factor can take between low and high levels.

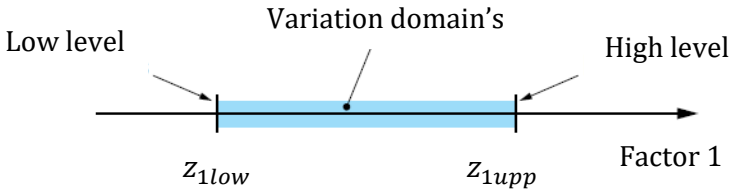


Figure I.2. Factor's domain of variation.

I.6.2. Response

The response in an experimental design is the output variable or variable being studied. It represents the measurable event observed when the factors under investigation are varied. The response should be representative, quantifiable, and exhibit minimal dispersion for controlled and constant input variables.

I.6.3. Study domain (Experimental domain)

It represents the part of the experimental space chosen by the experimenter to carry out the trials. It is defined by the combination of the domains of the different factors studied and delimited by their lower and upper levels. Figure I.3a shows a study domain defined by two factors, while Figure I.3b illustrates one comprising three factors, thus extending the experimental space (the blue area).

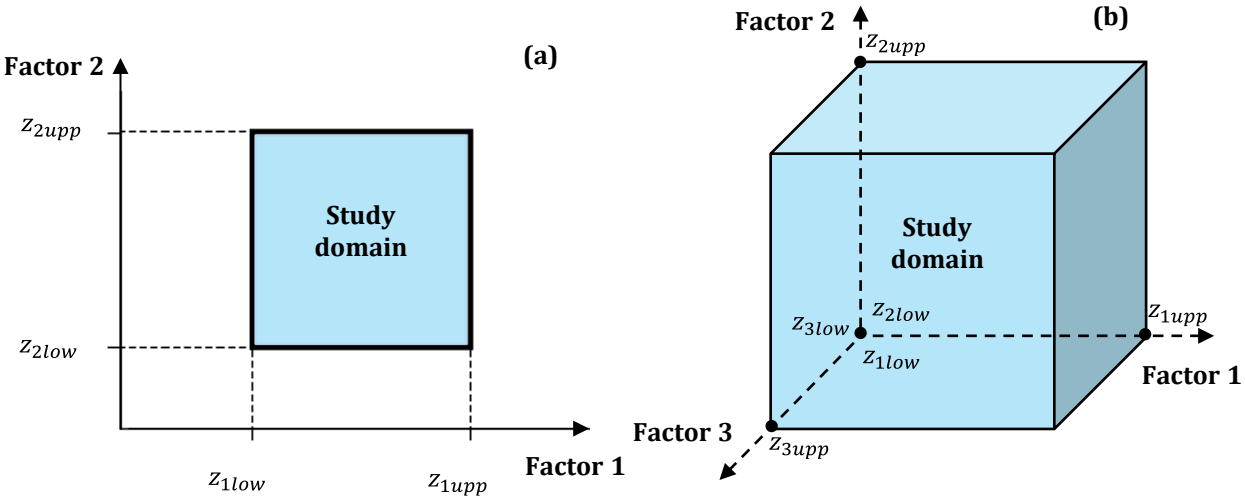


Figure I.3. Experimental space or study domain.

I.6.4. Response surface

Each point in the study domain, delimited by the levels of each factor, corresponds to a response. All the points in the study domain correspond to a set of responses which are located on a surface, thus defining the concept of a response surface (Figure I.4).

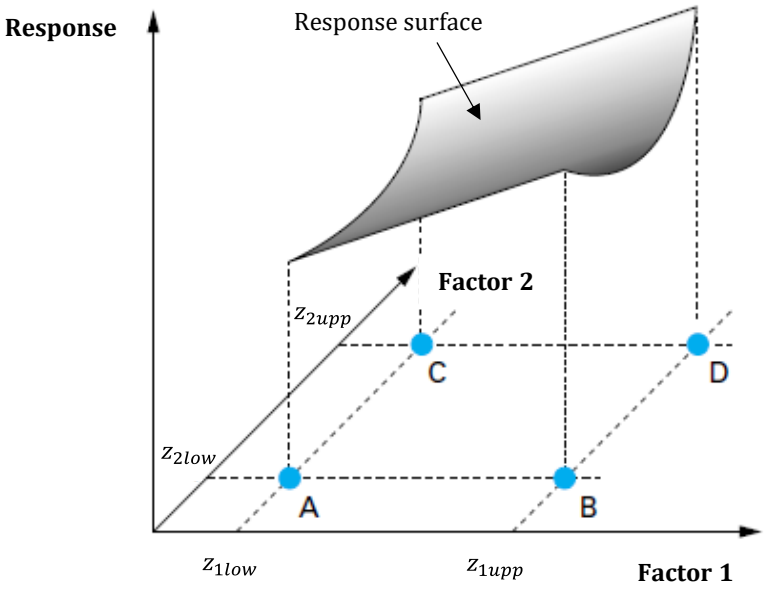


Figure I.4. Response surface presentation.

Plotting these surfaces helps find optimal operating conditions for maximizing or minimizing characteristics like yield, product purity, and cost through successive approximations.

I.6.5. Coded variables (centred variables)

Coded variables are dimensionless variables. They are used in postulated models to facilitate data processing. Each real variable $\boldsymbol{z_j}$ is associated with a coded variable $\boldsymbol{x_j}$. The use of coded variables makes it possible to associate the real variables $z_{j(low)}$ and $z_{j(upp)}$ with the normalized values -1 and +1, respectively.

$$z_{j(low)} \longrightarrow x_j = -1$$

$$z_{j(upp)} \longrightarrow x_j = +1$$

The transition from real variables $(z_1, z_2, \ldots\ldots z_k)$ to coded variables $(x_1, x_2, \ldots\ldots x_k)$, and vice versa, is given by the following coding formula :

$$x_j = \left(\frac{z_j - z_j^0}{\Delta z_j}\right), \quad j = 1,2, \ldots\ldots, k \qquad \boldsymbol{Eq.(I-1)}$$

with :

$$z_j^0 = \frac{z_{jupp} + z_{jlow}}{2} \qquad and \qquad \Delta z_j = \frac{z_{jupp} - z_{jlow}}{2}$$

z_j^0 : the real variable that represents the center of the study domain ;

Δz_j : variation step size.

Example I.1

A researcher aims to extract bezacryl yellow (BY) from waste shoe soles using batch reactor. The operating parameters selected for this study are pHϵ[2, 10], initial concentration of JB, i.e. $[BY]_0\epsilon[10, 50]$ mg/L, and dosage of shoe sole waste (SSW), i.e. SSW dosage ϵ[0.4, 4] g/L.

1. Calculate the upper and lower levels of the factors in terms of coded values.
2. Find the coded values corresponding to to real values of pH=8, $[BY]_0$= 45 mg/L and SSW dosage =1.5g/L.
3. What is the real pH if its coded value is estimated to be -0.7?

Resolution

1. **Determination of coded values**

Let's calculate the values at the centeral point of the investigating domains and the variation step size of the different factors:

pH : $z_j^0 = \frac{10+2}{2} = 6$; $\Delta z_j = \frac{10-2}{2} = 4$; $\begin{cases} z_{1upp} = 10 \rightarrow x_1 = \left(\frac{10-6}{4}\right) = +1 \\ z_{1low} = 2 \rightarrow x_1 = \left(\frac{2-6}{4}\right) = -1 \end{cases}$

$[BY]_0$: $z_j^0 = \frac{50+10}{2} = 30$; $\Delta z_j = \frac{50-10}{2} = 20$; $\begin{cases} z_{2upp} = 50 \rightarrow x_2 = \left(\frac{50-30}{20}\right) = +1 \\ z_{2low} = 10 \rightarrow x_2 = \left(\frac{10-30}{20}\right) = -1 \end{cases}$

SSW dosage : $z_j^0 = \frac{4+0.4}{2} = 2.2$; $\Delta z_j = \frac{4-0.4}{2} = 1.8$; $\begin{cases} z_{3upp} = 4 \rightarrow x_3 = \left(\frac{4-2.2}{1.8}\right) = +1 \\ z_{3low} = 0.4 \rightarrow x_3 = \left(\frac{0.4-2.2}{1.8}\right) = -1 \end{cases}$

All the calculations are summarised in the table below.

Table I.1. Natural factors, formula parameters and coded variables calculated from data in Example I.1.

Factors		**Formula's Parameters**		**Coded variables**
Natural variables	**Intervall of variation**	z_j^0	Δz_j	x_j
z_1: pH solution	[2, 10]	6	4	$x_1 \begin{cases} -1 \\ +1 \end{cases}$
z_2: $[BY]_0$ mg/L	[10, 50]	30	20	$x_2 \begin{cases} -1 \\ +1 \end{cases}$
z_3: SSW dosage g/L	[0.4, 4]	2.2	1.8	$x_3 \begin{cases} -1 \\ +1 \end{cases}$

2. **<u>The coded values corresponding to to real values of pH=8, $[BY]_0$= 30 mg/L and SSW dosage=2</u>**

Let's calculate the coded values of each factors:

pH = 8 : $x_1 = \left(\frac{8-6}{4}\right) = 0.5$

$[BY]_0$=45 mg/L: $x_2 = \left(\frac{45-30}{20}\right) = 0.75$

SSW dosage = 1.5 g/L: $x_3 = \left(\frac{1.5-2.2}{1.8}\right) = -0.39$

3. **<u>The real pH value when its coded value is estimated to be -0.7</u>**

Let's calculate the real value of pH using the folwing formula :

$z_j = \Delta z_j . x_j + z_j^0$ then $z_1 = \Delta z_1 . x_1 + z_1^0$

When $x_1 = -0.7$; $z_1 = 4.(-0.7) + 6 = 3.2$, then pH $= 3.2$

I.6.6. Experiment matrix

Experiment matrix is a table in which each row corresponds to an experiment or trial, and each column represents a factor. It is recommended for factors studied over three, as geometric representation becomes practically impossible in multidimensional spaces beyond three. The experiment matrix denoted T(N, k) lists all the experiments to be conducted, either in coded (+1, -1) or real form.

N : number of rows, representing the number of experiments to be carried out ;

K : number of columns, indicating the number of factors studied.

For example, a full factorial design study of pHϵ[2, 10] and [[BY]$_0\epsilon$[10, 50] mg/L necessitates a four-trial experiment matrix to examine a process involving two factors T(4,2) (see Chapitre II).

Table I.2. Experiment matrix of full factorial design at two factors T(4,2).

	Coded factors		**Real factors**	
N	x_1	x_2	z_1	z_2
1	-1	-1	2	10
2	+1	-1	2	50
3	-1	+1	10	10
4	+1	+1	10	50

I.6.7. Effects matrix

Effects matrix (X) is a tool utilized for calculating all coefficients in a postulated model equation through matrix calculation. It is constructed by adding, to the left of the matrix of experiments, a column composed solely of values (+1), representing the ficitve variable x_0. To the right of the matrix, columns are added corresponding to the various possible interactions between the factors. These interactions are determined using the mathematical model postulated for the study. The effects matrix is denoted: X (N, M).

M : number of columns, indicating the number of unknowns in the model (number of coefficients).

The effects matrix for the two-factor example provided previously is X(4, 4), consisting 4 rows and 4 columns.

Table I.3. Effects matrix of full factorial design at 2 factors X(4,4).

	Effects			
N	x_0	x_1	x_2	x_{12}
1	+1	-1	-1	+1
2	+1	+1	-1	-1
3	+1	-1	+1	-1
4	+1	+1	+1	+1

I.6.8. Factor effect

The impact of a factor x_j on response y is determined by comparing the values taken by x_j as it moves from a lower level (x_j= -1 → y_1) to a higher level ((x_j= +1 → y_2). This effect or impact can be main (average) : $(\frac{y_2-y_1}{2})$ or global : $(y_2 - y_1)$ (Figure I.5).

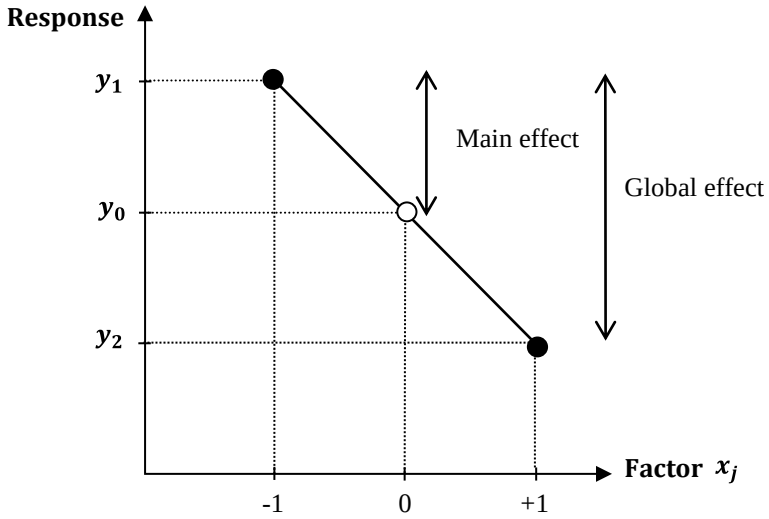


Figure I.5. Illustration of global effect and main effect.

It should be noted that the direction of variation indicates whether the factor has a positive or negative effect on the response. A positive design variable effect indicates a higher main response at a higher parameter setting, while a negative effect indicates a higher main response at a lower parameter setting (Figure I.6). However, the slope's steepness allows for the quick identification of the most influential factors.

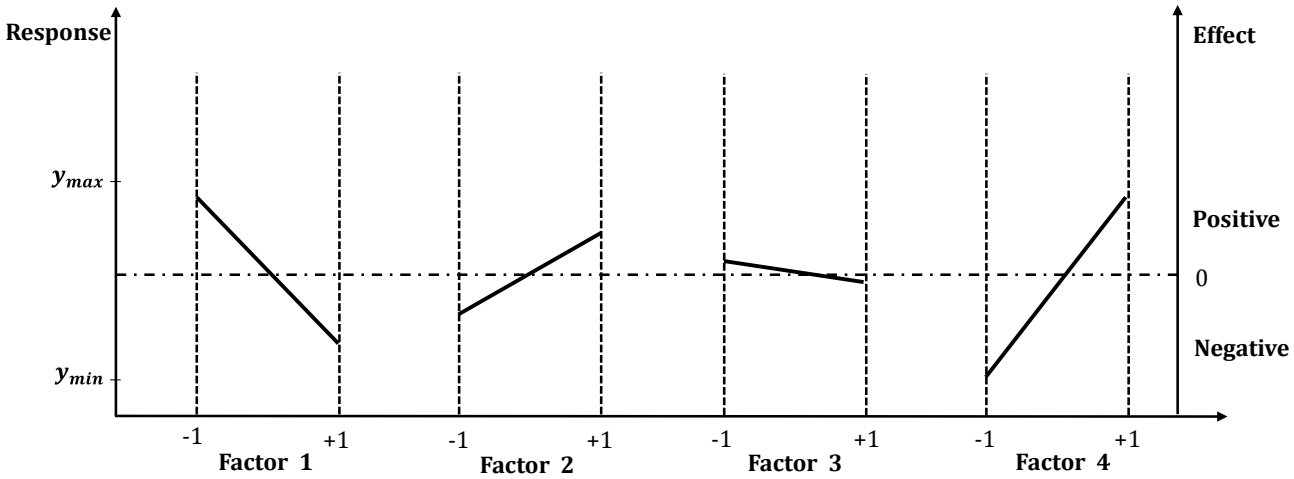


Figure I.6. Graphical variation of factors main effect.

I.6.9. Notion of interaction

Factors x_1 and x_2 interact if their effect on response y depends on x_2's level and vice versa. The effect a_{12} measures the variation in x_1's effect as x_2 increases from low to high level. Figure I.7 shows three types of interaction : (a) parallel lines, which indicate the abscence of interaction, and crossed lines, which indicate (b) low and (c) strong interactions.

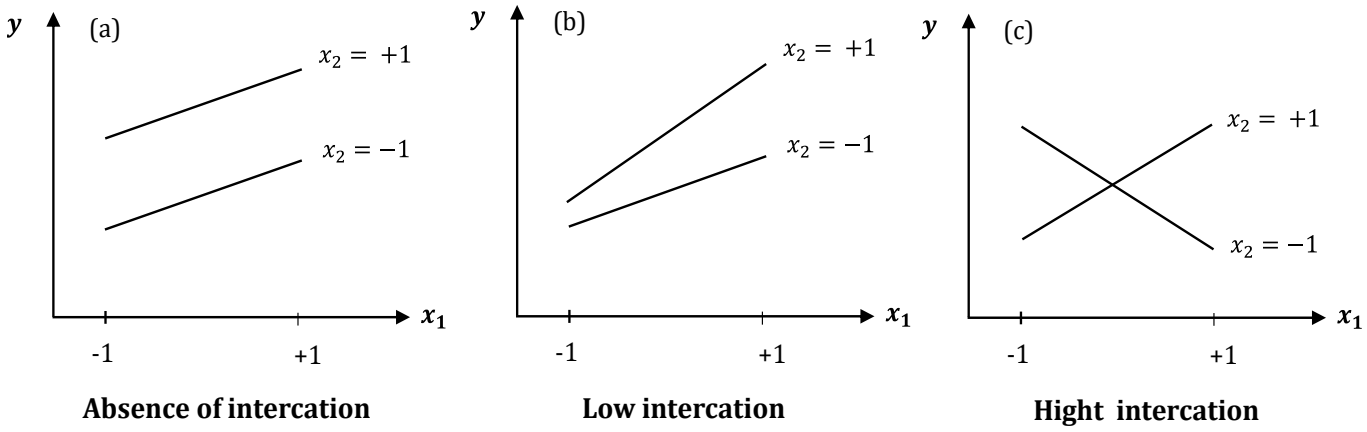


Figure I.7. Illustration of interactoion between x_1 and x_2.

As shown in Figure I.7(a) factors x_1 and x_2 are independent, as their respective effects do not depend on the levels of the other (parallel lines). Figure I.7(b) and (c), on the other hand, reveal an interaction between x_1 and x_2, with non-parallel lines. In Figure I.7(c), the inversion of the effect of x_1 according to the level of x_2 highlights a particularly strong interaction.

Chapter II
Full Factoriel Design at Tow Level 2k

Chapter II: Full Factoriel Design at Tow Level 2^k

II. 1. Introduction

Researchers are often interested in understanding the effects of multiple factors operating together. However, without a carefully thought-out and well-executed research design that considers all of those factors simultaneously, potentially misleading or biased estimates of the effects of the factors may be produced. The use of a full factorial design where the combination of all levels of the factors is thoroughly investigated can greatly enhance and improve the efficiency of statistical analysis as compared to other experimental methodologies. This approach offers a structured and rigorous study framework, making it possible to minimise experimental errors and optimise the reliability of the results obtained. It also provides a basis for the development of more advanced experimental designs, such as fractional factorial designs and response surface designs. These techniques build on the principles of full factorial design to efficiently explore complex systems and optimize processes.

II.2. Concept of modeling and multiple linear regression

Without prior knowledge of the functional relationship between the response and the factors, experimenters define an a priori evolutionary law as follow :

$$y = f(x_1, x_2, x_3, \ldots\ldots x_n) + \varepsilon$$

with :

y: experimental or measured response ;

ε : residue or experimental error ;

$f(x_1, x_2, x_3, \ldots\ldots x_n)$: unknown function which explains best the variations in response according to the different values given to factors.

This unknown function is too complex to model precisely without making simplifying assumptions. To approximate it, polynomial smoothing functions are often used. This technique leverages Taylor's theorem, which states that any smooth function can be locally approximated by a polynomial (Eq. II-1). By fitting a polynomial to the data, the underlying function's behavior within a specific region can be estimated.

$$\hat{y} = a_0 + \sum_{j=1}^{k} a_j x_j + \sum_{\substack{u,j=1 \\ j \neq u}}^{k} a_{ju}\, x_j x_u + \sum_{\substack{l,u,j=1 \\ j \neq u \neq l}}^{k} a_{jul\ldots k} x_j x_u x_l \ldots x_k + \sum_{j=1}^{k} a_{jj}\ x_j^2 \qquad (\boldsymbol{Eq.\,I-1})$$

with :

$\hat{y}$: response calculated from the mathematical postulated model ;

x_j : coded variables ;

k : number of factors studied ;

$a_0, a_j, a_{ju}, a_{jul\ldots k}, a_{jj}$: coefficients of the mathematical postulated model:

a_0 : constant ;

a_j : linear effects ;

$a_{ju}, a_{jul\ldots k}$: interaction effects ; ;

a_{jj} : quadratic effects.

Experimental designs employ the multilinear regression technique to determine the coefficients of the postulated polynomial model, based on the method of least squares, which minimizes the sum of the squared differences between the experimental and mathematical variables.

$$\varphi = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 = min$$

The advantage of modelling the measured response by a polynomial lies in the possibility of using all the results of matrix algebra. Although other mathematical functions can be used, experience has shown that polynomials solve most problems.

II.3. Full factorial design at two levels 2^k

A full factorial design 2^k (FFD) is a systematic experimental design technique that investigates the effects of multiple factors ***k*** on a response variable. It is widely used due to its simplicity and speed of implementation. FFD involves **N= 2^k** experiments, testing all possible combinations of factor at two levels (low level -1 and high level +1), allowing researchers to identify the main effects of each factor and their interactions.

In a full factorial design with three factors, each factor being at two levels, the experimental space is a cube with eight vertices (Figure II.1), i.e. eight experimental trials (N=2^3). Each vertex corresponds to an experiment and its coordinates, expressed as

coded variables, define the levels of the three factors for this experiment. These coordinates are represented in the experimental design matrix (Table II.1).

Figure II.1. Layout of experimental points in a full factorial design 2^3.

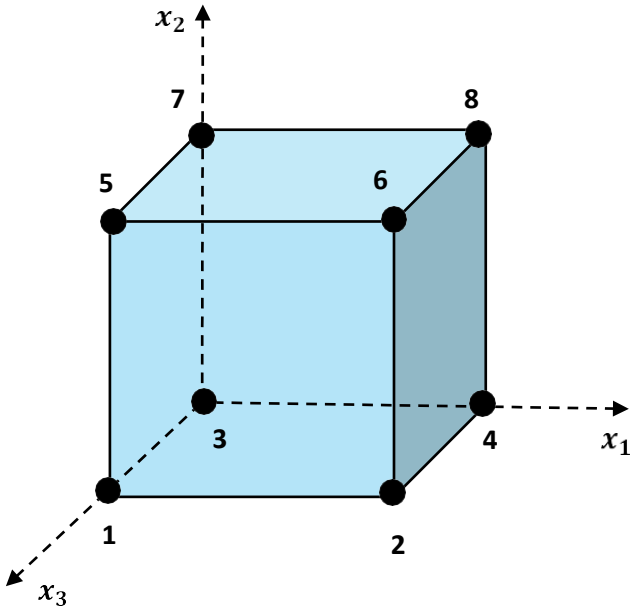


Table II.1. Experiment matrix (N, k)= (8,8) of a full factorial design 2^3.

Run	**Real factors**			**Coded factors**			**Response**
N	z_1	z_2	z_3	x_1	x_2	x_3	y_i
1	z_{1low}	z_{2low}	z_{3low}	-1	-1	-1	y_1
2	z_{1upp}	z_{2low}	z_{3low}	+1	-1	-1	y_2
3	z_{1low}	z_{2upp}	z_{3low}	-1	+1	-1	y_3
4	z_{1upp}	z_{2upp}	z_{3low}	+1	+1	-1	y_4
5	z_{1low}	z_{2low}	z_{3upp}	-1	-1	+1	y_5
6	z_{1upp}	z_{2low}	z_{3upp}	+1	-1	+1	y_6
7	z_{1low}	z_{2upp}	z_{3upp}	-1	+1	+1	y_7
8	z_{1upp}	z_{2upp}	z_{3upp}	+1	+1	+1	y_8

Each experimental condition is termed a **run**, and the associated response measurement is referred to as an **observation**. The entire collection of runs is known as the **design**.

II.4. Mathematical models associated to full factorial design

A full factorial design at two level can be used for both continuous and discrete variables.

The mathematical model associated with it is generally a first-order synergistic polynomial model, integrating all the factors studied. It is generally expressed in the following form:

$$\hat{y} = a_0 + \sum_{j=1}^{k} a_j x_j + \sum_{\substack{u,j=1 \\ j \neq u}}^{k} a_{ju}\, x_j x_u + \sum_{\substack{l,u,j=1 \\ j \neq u \neq l}}^{k} a_{jul\ldots k} x_j x_u x_l \ldots x_k \qquad (\boldsymbol{Eq.\,I-2})$$

or :

$$\hat{y} = a_0 + a_1 x_1 + a_2 x_2 + \cdots a_k x_k + a_{12} x_1 x_2 + a_{13} x_1 x_3 + \cdots + a_{1..k} x_1 .. x_k \qquad (\boldsymbol{Eq.\,I-3})$$

Depending on the number of factors ***k*** studied, the polynomial (Eq. II-3) can be formulated in a more or less complex way. It must essentially include 2^k coefficients, which are the unknowns (M) and correspond to:

- k main effects (linear) ;
- (2^k)-k-1 interactions ;
- a constant representing the mean response.

This polynomial model must clearly reflect the effect of each factor, whether it is individual (absolute effect) or the result of double, triple or even higher level interactions if they exist. The results of a full factorial design at two levels thus lead to a system of $\boldsymbol{N= 2^k}$ equations for $\boldsymbol{M= 2^k}$ unknowns, with. **N=M**.

II.5. Determination of model coefficients

The coefficients representing the linear effects and interactions in the polynomial model developed (Eq. II-3) are calculated by mathematically solving a system of 2k equations with 2k unknowns, using the following formula :

$$A = \frac{1}{N} X^T Y$$

with :

A : model coefficients vector ;

X^T : transpose of the effects matrix X ;

X : effects matrix (N, M) ;

Y : observations or responses vector.

However, it is preferable to solve the regression in matrix form, as this approach offers several advantages in terms of simplicity, efficiency and robustness. The matrix formulation enables simultaneous manipulation of all equations, making it beneficial for large systems and promoting the use of optimized numerical algorithms, thereby reducing calculation errors and problem solving time.

II.6. Regression analysis in matrix form

Regression analysis in matrix form is particularly well-suited to computerised resolution, offering an efficient method for dealing with complex problems. Using the method of least squares, it is possible to determine the coefficients of the regression equation (Eq. II-3).

Various statistical tools can be extracted from this equation, including :

1) Effects matrix X (N, M):

$$X = \begin{bmatrix} x_{01} & x_{11} & x_{21} & \dots & \dots & x_{M1} \\ x_{02} & x_{12} & x_{22} & \dots & \dots & x_{M2} \\ \vdots & \vdots & \vdots & \dots & \dots & \vdots \\ \vdots & \vdots & \vdots & \dots & \dots & \vdots \\ x_{0N} & x_{1N} & x_{2N} & \dots & \dots & x_{MN} \end{bmatrix}$$

2) Model coefficients vector A (M, 1) :

$$A = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ \vdots \\ a_{1..k} \end{bmatrix}$$

3) Vecteur des observations Y *(N, 1)* :

$$Y = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ \vdots \\ y_N \end{bmatrix}$$

Thus, the postulated mathematical model (Eq. I-3) can be written in the following matrix form :

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} x_{01} & x_{11} & x_{21} & \cdots & \ldots & x_{M1} \\ x_{02} & x_{12} & x_{22} & \ldots & \ldots & x_{M2} \\ \vdots & \vdots & \vdots & \ldots & \ldots & \vdots \\ \vdots & \vdots & \vdots & \ldots & \ldots & \vdots \\ x_{0N} & x_{1N} & x_{2N} & \ldots & \ldots & x_{MN} \end{bmatrix} \times \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ \vdots \\ a_{1..k} \end{bmatrix} \qquad (\boldsymbol{Eq.\,I-4})$$

By introducing the transpose of the effects matrix X; i.e X^T (M, N); to (Eq. II-4), we can have a system of normal equations designed to determine $a_0, a_j, a_{ju}, a_{jul\ldots k}$.

$$\sum_{i=1}^{N} x_{0i}y_i = a_0 \sum_{i=1}^{N} x_{0i}^2 + a_1 \sum_{i=1}^{N} x_{0i}x_{1i} + \cdots + a_M \sum_{i=1}^{N} x_{0i}x_{Mi}$$

$$\sum_{i=1}^{N} x_{1i}y_i = a_0 \sum_{i=1}^{N} x_{1i}x_{0i} + a_1 \sum_{i=1}^{N} x_{1i}^2 + \cdots + a_M \sum_{i=1}^{N} x_{1i}x_{Mi}$$

……

$$\sum_{i=1}^{N} x_{Mi}y_i = a_0 \sum_{i=1}^{N} x_{Mi}x_{0i} + a_1 \sum_{i=1}^{N} x_{Mi}x_{1i} + \cdots + a_M \sum_{i=1}^{N} x_{Mi}^2$$

All these equations can be written in the following matrix form :

$$X^T Y = [X^T X] \quad A$$

with :

$$X^T = \begin{bmatrix} x_{01} & x_{02} & \ldots & \ldots & x_{0N} \\ x_{11} & x_{12} & \ldots & \ldots & x_{1N} \\ x_{21} & x_{22} & \ldots & \ldots & x_{2N} \\ \vdots & \vdots & \cdots & \ldots & \vdots \\ \vdots & \vdots & \ldots & \ldots & \vdots \\ x_{M1} & x_{M2} & \ldots & \ldots & x_{MN} \end{bmatrix}$$

X^T (M, N) : is obtained by exchanging the rows and columns of X (N,M).

$[X^T X]$: matrix of variances.

$$[X^T X] = \begin{bmatrix} \sum x_{0i}^2 & \sum x_{0i}x_{1i} & \ldots & \ldots & \sum x_{0i}x_{Mi} \\ \sum x_{1i}x_{0i} & \sum x_{1i}^2 & \ldots & \ldots & \sum x_{1i}x_{Mi} \\ \vdots & \vdots & \ldots & \ldots & \vdots \\ \vdots & \vdots & \ldots & \ldots & \vdots \\ \sum x_{Mi}x_{0i} & \sum x_{Mi}x_{1i} & \ldots & \ldots & \sum x_{Mi}^2 \end{bmatrix}$$

$[X^T Y]$: column matrix.

$$[X^T Y] = \begin{bmatrix} \sum x_{0i} y_i \\ \sum x_{1i} y_i \\ \vdots \\ \vdots \\ \sum x_{Mi} y_i \end{bmatrix}$$

The vector of coefficients, A, is obtained from the product :

$$A = [X^T X]^{-1} X^T Y \qquad (\boldsymbol{Eq.\,I-5})$$

where : $[X^T X]^{-1}$ is the inverse matrix* of the matrix $[X^T X]$. In the case of the full factorial design, the variance-covariance matrix $[X^T X]^{-1}$ is diagonal and is represented as follows:

$$[X^T X]^{-1} = \begin{bmatrix} 1/N & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/N \end{bmatrix}$$

II.7. Optimality of experimental designs

An optimal experimental design ensures the calculation of coefficients of the effects of factors (a_j) with the best accuracy, minimizing model response error. To achieve this, several criteria must be met, including the most commonly used ones are :

II.6.1. Unit matrix criterion (Hadamard matrix)

To achieve the minimum variance in N experiments, the square matrix of effects X, consisting of +1 and -1 elements (Hadamard matrix), associated with factorial designs must satisfy the following relationship :

$$[X^T X] = N[I], \; I \text{ : unit matrix.}$$

The unit matrix criterion ensures that :

- the factors are tested in a balanced and independent manner ;
- the estimates of the effects are not biased by correlations between the columns of the matrix.

---

* : a square matrix A is invertible if there is an inverse matrix $\mathbf{A^{-1}}$ such that $\mathbf{A \times A^{-1} = A^{-1} \times A = I}$, and only square matrices can have inverses.

II.7.2. Orthogonality criterion

The effects matrix X is orthogonal if its column vectors are orthogonal in pairs, meaning the scalar product of two columns is zero.

Let there be two column vectorsX_1 and X_2 of the effects matrix X :

$$X_1 = \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1i} \\ \vdots \\ x_{1N} \end{bmatrix} \qquad and \qquad X_2 = \begin{bmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2i} \\ \vdots \\ x_{2N} \end{bmatrix}$$

The orthogonality of these two columns is expressed as:

$$\sum_{i=1}^{N} x_{1i} \times x_{2i} = 0$$

II.7.3. Minimum trace criterion

The $[X^T X]^{-1}$ matrix's trace, representing the sum of its main diagonal elements, must beas small as possible to minimize error, resulting in :

$$tr([X^T X]^{-1}) = \sum_{i=1}^{N} ([X^T X]^{-1})_{ii} = minimum\ possible$$

II.8. Example of application

II.8.1. Problem statement

A manufacturer wants to recover the amoxicillin molecules (AMX) present in his industrial effluents using an adsorption process based on a mineral adsorbent (hydroxyaptite: HAP). He wanted to study the influence of certain parameters like the pH of the solution (pH∈[2-10]), the AMX initial concentration ([AMX]∈[50-300] mg/L) and the HAP dosage ([HAP]∈[0.125-1.25] g/250mL) on the adsorption efficiency of AMX onto HAP. To do this, he decided to adapt a full factorial design at two levels (FFD).

II.8.2. Problem formulation

- Objective of the study : study the effect of each operating parameter on the rate ofadsorption of AMX.
- Define the factors and the response :

Response y: adsorption rate calculated as follows:

$$y\,(\%) = \frac{[AMX]_0 - [AMX]_t}{[AMX]_0} \times 100$$

with:

$[AMX]_0$ and $[AMX]_t$: are initial concentration of amoxicillin and its concentration at time t.

Factors and their study domain :

Factor 1 ⇒z_1: pH of the solution to be treated; pH∈[2-10] ;

Factor 2 ⇒z_2: AMX concentration ; [AMX]∈[50-300]mg/L ;

Factor 3 ⇒z_3: PAH concentration in the suspension ; [HAP]∈[0.125-1.25]g/250mL.

A study using a full factoriel design at two-level and three-factor requires N = 2^3 = 8 experiments, with a mathematical model with M = 8 unknown coefficients, as per the given form :

$$\hat{y} = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{23}x_2x_3 + a_{123}x_1x_2x_3$$

II.8.3. Planning and experiments

- construction of the experimental design: construct the experiment matrix in real and coded variables (Table II.2), which serves as a start-up tool for launching and facilitating the experiment's implementation

Table II.2. Experiment matrix of full factorial design at two factors 2^3.

Run N	Real factors			Coded factors			Response observed
	z_1	z_2	z_3	x_1	x_2	x_3	$y(\%)$
1	2	50	0.125	-1	-1	-1	55.89
2	10	50	0.125	+1	-1	-1	56.93
3	2	300	0.125	-1	+1	-1	64.70
4	10	300	0.125	+1	+1	-1	61.23
5	2	50	1.25	-1	-1	+1	85.95
6	10	50	1.25	+1	-1	+1	89.95
7	2	300	1.25	-1	+1	+1	88.50

8	10	300	1.25	+1	+1	+1	96.55

II.8.4. Results analysis

To maximize information from experimental results, calculate the coefficients of the postulated model (a_j), i.e. A vector, by determining the effects matrix X (N, M) (Tbele II.3). The vector A is calculated using the formula: $A = [X^T X]^{-1} X^T Y$, with the results listed in Table II.4.

Table II.3. Effects matrix X (8, 8)

N°d'essai	x_0	x_1	x_2	x_3	x_{12}	x_{13}	x_{23}	x_{123}	*y(%)*
1	+1	-1	-1	-1	+1	+1	+1	-1	55.89
2	+1	+1	-1	-1	-1	-1	+1	+1	56.93
3	+1	-1	+1	-1	-1	+1	-1	+1	64.70
4	+1	+1	+1	-1	+1	-1	-1	-1	61.23
5	+1	-1	-1	+1	+1	-1	-1	+1	85.95
6	+1	+1	-1	+1	-1	+1	-1	-1	89.95
7	+1	-1	+1	+1	-1	-1	+1	-1	88.50
8	+1	+1	+1	+1	+1	+1	+1	+1	96.55

Table II.4. Coefficient values for the AMX adsorption model.

Constant	**Linear effects**			**Iinteraction effects**			
a_0	a_1	a_2	a_3	a_{12}	a_{13}	a_{23}	a_{123}
74.96	1.20	2.78	15.28	-0.06	1.81	-0.49	1.07

The model obtained after calculation is then written :

$$\hat{y}\ (\%) = 74.96 + 1.20x_1 + 2.78x_2 + 15.28x_3 - 0.06x_1x_2 + 1.81x_1x_3 - 0.49x_2x_3 + 1.07x_1x_2x_3$$

This postulated model enables the differentiation of the three operating parameters studied on the rate of adsorption of AMX by PAHs, allowing for the determination of each factor's influence in absolute terms and the identification of the combined influence of two factors simultaneously.

The final stage of the experimental design chronology, which is the optimisation of the response, can only be carried out once the postulated model has been validated by the static tests (Chapter III).

Chapter III
Statistical Analysis : Significance testing and model validation

Chapter III : Statistical Analysis : Significance testing and model validation

III.1. Introduction

The use of regression and correlation analysis allows for the identification of dependence relationships between factors (variables) and optimal conditions in experimental data. After determining the regression equation, the results are statistically analyzed. To successfully conduct this analysis, the following conditions must be met:

- the parameter $\boldsymbol{x}$ is measured with negligible error. The observed error in the response $\boldsymbol{y}$ can be explained by the presence of variables not taken into account in the process, which do not appear in the regression equation (Non-controlled and uncontrolabe factors) ;
- the observations on the output variables $\boldsymbol{y_1}, \boldsymbol{y_2}$ ... ... are independent and follow a normal distribution.

III. 2. Verification of the significance of the coefficients (effects)

II.2.1. Each test is repeated *m* times

When the experiments are repeated $\boldsymbol{m}$ times at the limits of the variation interval for each factor (level -1 and level +1), this following analysis is applied :

The arithmetic mean of the parallel test results $\overline{\boldsymbol{y_i}}$ is determined **[2, 11, 12]** :

$$\overline{y_i} = \frac{\sum_{u=1}^{m} y_{iu}}{m} \qquad (\boldsymbol{Eq.III-1})$$

then the sampling variance $\boldsymbol{S_i^2}$ for trial (experiment) $\boldsymbol{i}$:

$$S_i^2 = \frac{\sum_{u=1}^{m}(y_{iu} - \overline{y_i})^2}{m-1}, \quad i = 1,2,\ldots.N \qquad (\boldsymbol{Eq.III-2})$$

with :

m : number of replicates for each trial ;

N : number of experiments;

y_i: experimental responses (observed/measured responses).

The reproducibility variance is calculated if the sampling variances are homogeneous. that is presentes as :

$$S_{rep}^2 = \frac{\sum_{i=1}^{N} S_i^2}{N} \qquad (\boldsymbol{Eq.III-3})$$

III.2.2. Central test is repeated n_o times

When the experiments are repeated $\boldsymbol{n_o}$ at the centre of the experimental domain (level 0), the reproducibility variance is estimated by the variance calculated at the centre of the experimental domain.

$$S_{rep}^2 = \frac{\sum_{i=1}^{n_0} (y_{i0} - \bar{y}_0)^2}{n_0 - 1} \qquad (\boldsymbol{Eq.III-4})$$

with :

$f = n_0 - 1$: degrees of freedom ;

n_0 : number of repetitions in the centre of the experimental domain ;

y_{i0} : measured responses at the centre of the expérimental domain

$\bar{y}_0$: arithmetic mean of the measurements taken at the centre of the domain, calculated by:

$$\bar{y}_0 = \frac{\sum_{i=1}^{n_0} y_{i0}}{n_0} \qquad (\boldsymbol{Eq.III-5})$$

The reproducibility variance is crucial for estimating the significance of coefficients in regression equations using Student's *t*-ratio. A coefficient is said to be significant if, for a given risk, it is significantly different from zero. In principle, two hypotheses are set up :

H_0 = « b_j = 0 »

against the hypothesis : H_1 = « $b_j \neq 0$ ».

If we accept the null hypothesis H_0 we automatically reject the alternative hypothesis H_1.

To do this, we calculate the Student's t_j-ratio :

$$t_j = \frac{|b_j|}{S_{bj}} \qquad (\boldsymbol{Eq.III-6})$$

where :

b_j: j^{ème} coefficient of the regression equation ;

s_{bj} : mean square deviation or standard deviation which, for a first-degree model is defined as follow:

$$S_{bj} = \sqrt{\frac{S_{rep}^2}{N}} \qquad (\boldsymbol{Eq.III-7})$$

The theoretical value of $\boldsymbol{t(p,f)}$ is determined using Student's table (Table III. A in the chapter Appendix) for the chosen significance level $\boldsymbol{p}$ and the number of degrees of freedom $\boldsymbol{f=n_0-1}$. The rule for this test is defined as follows :

- if $t_j > t_p(f)$, H_0 is rejected at the accepted risk ;
- if $t_j < t_p(f)$, H_0 is accepted at the accepted risk.

The H_0 hypothesis accepted states that the coefficient at the $\boldsymbol{p}$ risk is not significantly different from zero, indicating that the associated variable has no influence on the response.

The software used for experimental design, such as Statistica, JMP, Minitab or Design-Expert, is based on these assumptions and assesses the significance of the coefficients by calculating the probability *p*-value*. A p-value close to zero (p <0.05) indicates the coefficient's influence, while a *p*-value close to one (p >0.05) indicates the coefficient's inability to be distinguished from zero and thus not influential.

Example III.1

When examining the impact of a significant factor on a response at significance level of 0.05, the null hypothesis H_0 suggests no significant effect, while the alternative hypothesis H_1 suggests a significant effect. So, in the results analysis, when we have for exexample:

- a *p*-value of 0.03 (p<0.05) : it means that there is a 3% probability of observing an effect as extreme as that measured, if the factor really had no effect at all. As conclusion, we could reject the null hypothesis (H_0) and conclude that the factor is significant (accept H_1).
- a *p*-value of 0.12 ((p>0.05) : it means that there is a 12% probability of obtaining the observed data (or more extreme data) if the null hypothesis is true. As conclusion, we could accept the null hypothesis (H_0) and conclude that the factor is not significant (reject H_1).

* : the p-value is a statistical measure used to assess the strength of the evidence against a null hypothesis (H_0) in a statistical test. It represents the probability of obtaining a result at least as extreme as the one observed, under the hypothesis that H_0 is true.

Remark

A low *p*-value does not prove that the alternative hypothesis is true, only that there is sufficient evidence to reject H_0.

III. 3. Model validation

The model validation tests described in the following sections can be applied to the equations of models established for the two cases previously mentioned: (1) when the repetitions are conducted in the experimental domain's center, and (2) when all the trials are repeated $\boldsymbol{m}$ times. To adapt the formulas to the second case, simply replace $\boldsymbol{y_i}$ with $\boldsymbol{\bar{y}_i}$, which corresponds to the mean of each experiment's observations.

III. 3.1. Lack of fit

The lack of fit in experimental designs is a measure of the inability of a statistical model to correctly describe the relationships between the factors studied and the observed response. This indicates that the mathematical model chosen does not fully capture the structure of the underlying phenomenon. The lack of fit occurs when :

- the model used is incorrect or incomplete (for example, important interactions or quadratic terms are omitted) ;
- the assumptions about the relationship between variables (linear, quadratic, etc.) do not correspond to reality.

The lack of fit compares the variation not explained by the model (residual variance $S^2_{rés}$) with the variation inherent in the experimental errors (reproducibility variance S^2_{rep}) :

$$F_{cal} = \frac{S^2_{rés}}{S^2_{rep}} \qquad (\boldsymbol{Eq.III-8})$$

The residual variance $S^2_{rés}$ is estimated as follows :

$$S^2_{rés} = \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N - \ell} \qquad (\boldsymbol{Eq.III-9})$$

with :

$N - \ell$: degrees of freedom ;

ℓ : number of significant coefficients ;

$\hat{y}_i$: responses calculated from the model (prédicted values).

If the following inequality holds true :

$$F_{cal} < F_{0.95}(N-\ell, n_0-1) \qquad (\boldsymbol{Eq.III-10})$$

$F_{0.95}(N-\ell, n_0-1)$: tabulated value of Fisher-Snedecor test (Table III. B in the chapter Appendix) for the chosen significance level $p(0.95)$ and the number of degrees of freedom $f_1 = N-\ell$ and $f_2 = n_0 - 1$.

this confirms that the model chosen is consistent and that it can be used to interpret the results or make predictions.

In the context of *p*-value analysis associated with the F-test in software used for experimental design, a lack-of-fit test producing a high *p*-value (for example, $p>0.05$) indicates that the statistical model fits the experimental data well. In other words, there is no statistically significant evidence to reject the null hypothesis that the proposed model is adequate to describe the relationship between the factors studied and the observed response.

III.3.2. Regression verification (Analyse of variance : ANOVA)

Checking the significance of the regression in the experimental designs means assessing whether the regression model obtained is globally relevant in explaining the variation in the experimental data. The Fisher test (F-test) is used to validate this analysis ; it is given by the following equation :

$$F_{cal} = \frac{\sum_{i=1}^{N}(\hat{y}_i-\bar{y})^2 / (\ell-1)}{\sum_{i=1}^{N}(y_i-\hat{y}_i)^2 / (N-\ell)} \qquad (\boldsymbol{Eq.III-11})$$

with :

$$\bar{y} = \frac{\sum_{i=1}^{N} y_i}{N} \qquad (\boldsymbol{Eq.III-12})$$

If the ratio F_{cal} is greater than the tabulated value $F_{0,95}(\ell-1, N-\ell)$ (Table III. B in the chapter Appendix) for the chosen significance level $p(0.95)$ and the numbers of degrees of freedom $f_1 = \ell - 1$ and $f_2 = N-\ell$, the equation is adequate. This F-test indicates how much the variance relative to the equation obtained is reduced in comparison with the variance relative to the mean. The efficiency of the regression equation is increased when the value of F_{cal} is greater than that of $F_{0.95}(\ell-1, N-\ell)$.

In the context of *p*-value analysis associated with the F-test in software used for experimental design, the regression verification or analyse of variance (ANOVA) tests the null hypothesis (H_0), which states that all the regression coefficients (except the y-intercept) are equal to zero, against the alternative hypothesis (H_1), which asserts that all the regression coefficients are significantes. A low p-value (p>0.05) indicates that the overall model is significant, i.e. that at least one of the coefficients is different from zero.

II.3.3. Coefficient of determination

The coefficient of determination R^2 is a statistical measure used to assess the goodness of fit of a model to a set of data, particularly in experimental design. It quantifies the effectiveness of the model by showing the extent to which observed variations in the response (dependent variable) are linked to variations in the factors manipulated in the experiment. R^2 is calculated as follows :

$$R^2 = \frac{\sum_{i=1}^{N}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{N}(y_i-\bar{y})^2} \qquad (\boldsymbol{Eq.III-13})$$

Depending on the value of R^2 obtained, several cases can be observed :

- $\mathbf{R^2 = 1}$ **:** this means that the model perfectly explains the variance in the data, i.e. all the variations in the response are explained by the factors in the model ;
- $\mathbf{R^2 = 0}$ **:** this means that the model does not explain any variance in the data, so the factors in the model are not related to the response at all ;
- $\mathbf{0 \leq R^2 \leq 1}$ **:** this means that the model explains a partial proportion of the variance in the data. The closer R^2 is to 1, the better the model is at explaining variations in the dependent variable.

Elsewhere, a higher R^2 value does not necessarily indicate a good fit or a good regression model, as adding a new variable (either the variable is significant or not) can increase the R^2 value, leading to poor prediction. To address this issue, an adjusted R^2 ($\bar{R}^2$) is introduced :

$$\bar{R}^2 = R^2 - (1 - R^2)\frac{\ell-1}{N-\ell} \qquad (\boldsymbol{Eq.III-14})$$

$\bar{R}^2$ is a more robust measure for assessing the quality of a regression model. It is used to correct the biases associated with adding variables to a statistical model. Furthermore, $\bar{R}^2$ takes into account both the quality of the fit and the number of variables in the model.

Likewise, it is essential to note that if :

- $\bar{R}^2$ is close to R^2, this indicates that the model is relatively simple and that the addition of variables has not introduced any over-fitting ;
- $\bar{R}^2$ is much lower than R^2, this suggests that the model includes insignificant variables. The addition of these variables has increased the complexity of the model without substantially improving its ability to explain the variance of the dependent variable.

III.3.4. Residuals analysis

An analysis of the residuals in the experimental designs is an essential step in assessing the quality and validity of the statistical model fitted to the experimental data. The residuals are the differences between the observed values and the values predicted by the model (Eq. III-15). Analysing these residuals enables us to check whether the model's assumptions are respected and to identify any problems that could invalidate the conclusions.

$$\varepsilon = y_i - \hat{y}_i \qquad (\boldsymbol{Eq.III-15})$$

There are several types of graphs and statistical tests used to analyse residuals, such as the residuals diagram (graph of residuals vs. predicted values, histogram or residual density, etc.) or the residuals normality test (Kolmogorov-Smirnov, etc.).

A good analysis of the residuals will confirm that the model's errors are random and not systematic, and that the factors studied do indeed influence the dependent variable without unobserved or biased effects.

III.4. Example of application

The validity of the statistical analysis presented in this chapter is checked by revisiting the example from Chapter II, paragraph II.8.1: Adsorption of AMX on HAP.

The postulated model obtained after calculation is :

$$\hat{y}\ (\%) = 74.96 + 1.20x_1 + 2.78x_2 + 15.28x_3 - 0.06x_1x_2 + 1.81x_1x_3 - 0.49x_2x_3 + 1.07x_1x_2x_3$$

III.4.1. Statistical analysis of the regression equation

To estimate the reproducibility variance, the test carried out in the centre of the experimental domain was repeated four times. The AMX adsorption efficiency for each of these tests are given in Table III.1.

Table III.1. AMX adsorption efficiency at the centre of the experimental domain.

$\mathbf{n_0}$	**1**	**2**	**3**	**4**	$\bar{\mathbf{y}}_{\mathbf{0}}(\%)$
$\boldsymbol{y_0}(\%)$	97.78	97.75	97.74	97.99	97.81

III.4.1.1. Significance of the coefficients of the regression equation

The Student's *t*-test is used to determine whether there are any non-significant coefficients among the model's coefficients, which will be eliminated from the equation because their impact on the AMX adsorption efficiency is insignificant. To do this, we determine the reproducibility variance and the mean standard deviation of the tests :

- $S^2_{rep} = 3.757.10^{-3}$;
- $S^2_{bj} = 2.167.10^{-2}$.

then we calculate the *tj* corresponding to each effect. The results obtained are listed in Table III.2.

TableI II.2. *tj* values.

Constant	**Linear effects**			**Interaction effects**			
t_0	t_1	t_2	t_3	t_{12}	t_{13}	t_{23}	t_{123}
3458,825	55,423	128,375	704,820	2,681	83,508	22,823	49,369

For the significance level p=0.05 and the number of degrees of freedom f = n_0 -1 = 3, the tabulated value of the bilateral Student's test (Student's Table) is equal to: $t_p(f) = t_{0.05}(3) = 3{,}18$. As the calculated value of $\boldsymbol{t_{12}}$ is less than the tabulated value, the corresponding coefficient $\boldsymbol{a_{12}}$ is not significant, which means that it is eliminated from the model equation.

III.4.1.2. Model validation

- **Lack of fit**

To confirm the validity of the postulated model, we calculated the representative residual variance (S_{res}^{2} =1.70.10^{-2}) and compared it with the reproducibility variance (S_{rep}^{2}) using the Fisher-Snedecor test : $F_{cal} = \frac{S_{rés}^{2}}{S_{rep}^{2}}$

As the calculated value of the F-test ($F_{cal} = 7.18$) is lower than the tabulated value for the p=0.05 significance level, i.e. $F_{0;95}(N - \ell, n_0 - 1) = F_{0.95}(1, 3) = 10.1$ (Fisher table), we can confirm that the model representing the adsorption efficiency of AMX onto HAP is not biased.

- **Significance test of the regression**

The model is unbiased, so we can check the significance of the regression using the Fisher

$$F_{cal}\text{test} : = \frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2 / (\ell - 1)}{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 / (N - \ell)}$$

The calculated F-test value is 12201.817, and the tabulated $F_{0.95}(N - \ell, n_0 - 1) = F_{0.95}(6, 1)$ value is 234. As $F_{cal} > F_{0.95}$, this indicates that the postulated regression equation is adequate and valid at 95%. Then, the equation for this model is written :

$$\hat{y}\ (\%) = 74.96 + 1.20x_1 + 2.78x_2 + 15.28x_3 + 1.81x_1x_3 - 0.49x_2x_3 + 1.07x_1x_2x_3$$

II.4.1.3. Reliability assessment

- **Coefficient of determination**

The postulated model's responses were evaluated against the measured responses using the coefficient of determination, which is of 0.999. Its value is practically close to 1, indicating that all variations in response within the experimental domain are explained by the postulated regression equation.

- **Residuals analysis**

The residuals analysis of the of the predicted model is shown in Figure III.1. As seen, the diagram reveals no relationship between the predicted values $\hat{\boldsymbol{y}}_{\boldsymbol{i}}$ and the residuals $\boldsymbol{\varepsilon}$, as the points appear to be arranged randomly. This result also reflects the absence of information in the residuals which allows us to assert that all response variations are explained by the chosen regression model.

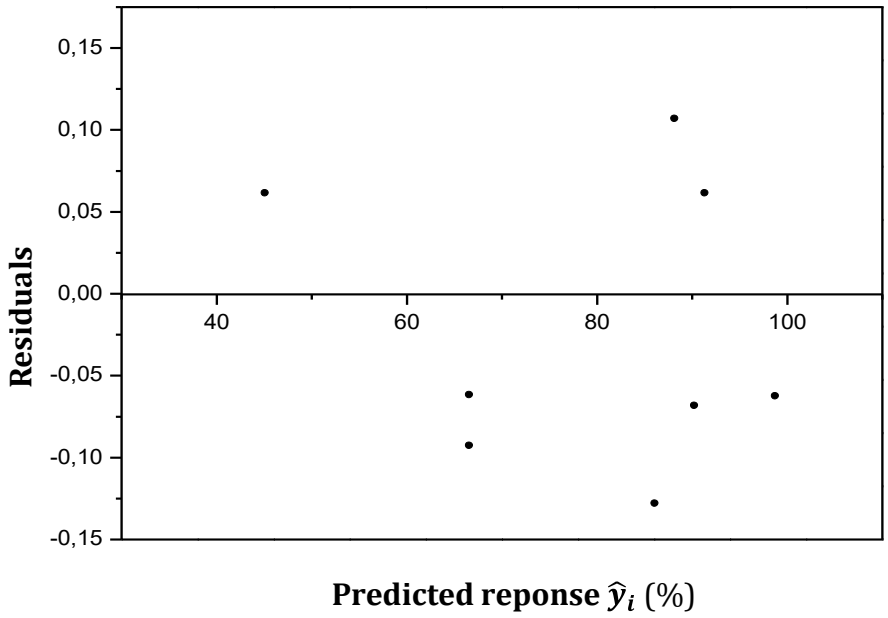


Figure III.1. Residuals diagram.

Table III. A. Student's law table

Unilatéral Bilatéral	**0,01** **0,20**	**0,05** **0,10**	**0,025** **0,05**	**0,01** **0,02**	**0,005** **0,01**
f_1					
1	3,08	6,31	12,7	31,8	63,7
2	1,89	2,92	4,30	6,97	9,92
3	1,64	2,35	3,18	4,54	5,84
4	1,53	2,13	2,78	3,75	4,60
5	1,48	2,02	2,57	3,37	4,03
6	1,44	1,94	2,45	3,14	3,71
7	1,42	1,90	2,37	3,00	3,50
8	1,40	1,86	2,31	2,90	3,36
9	1,38	1,83	2,26	2,82	3,25
10	1,37	1,81	2,23	2,76	3,17
11	1,36	1,80	2,20	2,72	3,10
12	1,36	1,78	2,18	2,68	3,06
13	1,35	1,77	2,16	2,65	3,01
14	1,35	1,76	2,15	2,62	2,98
15	1,34	1,75	2,13	2,60	2,95
16	1,34	1,75	2,12	2,58	2,92
17	1,33	1,74	2,11	2,57	2,90
18	1,33	1,73	2,10	2,55	2,88
19	1,33	1,73	2,09	2,54	2,86
20	1,33	1,73	2,09	2,53	2,85
21	1,32	1,72	2,08	2,52	2,83
22	1,32	1,72	2,07	2,51	2,82
23	1,32	1,71	2,07	2,50	2,81
24	1,32	1,71	2,06	2,49	2,80
25	1,32	1,71	2,06	2,49	2,79
26	1,32	1,71	2,06	2,48	2,78
27	1,31	1,70	2,05	2,47	2,77
28	1,31	1,70	2,05	2,47	2,76
29	1,31	1,70	2,05	2,46	2,76
30	1,31	1,70	2,04	2,46	2,75
40	1,30	1,68	2,02	2,42	2,70
∞	1,28	1,65	1,96	2,33	2,58

Table III. B. Table of Fisher-Senedecor distribution for p = 0.95

f_1 : degrees of freedom in numerator ;

f_2 : degrees of freedom in the denominator.

$f_1 \rightarrow$ $f_2 \downarrow$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254
2	18,5	19,00	19,2	19,20	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5	19,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,7	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6.00	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,37
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,3	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,8	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,7	2,65	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,19	2,15	2,1	2,06	2,01	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,07	1,99	1,95	1,9	1,85	1,8	1,75	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,2	2,13	2,06	1,97	1,93	1,88	1,84	1,79	1,73	1,67
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,71	1,65
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,03	1,94	1,90	1,85	1,81	1,75	1,70	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35	1,25
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22	1,00

Chapter IV
Fractional factoriel designs

Chapter IV : Fractional factoriel designs

IV.1. Introduction

In a full factorial design, the number of trials increases exponentially with the number of factors, rapidly making the experimental process impractical due to its complexity and high costs. For example, a full factorial design at two levels and six factors would require $2^6 = 64$ trials, while a ten factors design would require $2^{10} = 1024$ trials. As k increases, this inflation makes it impractical to carry out full experiments, particularly in cost- or time-constrained environments. Therefore, fractional factorial designs have been developed to compensate for this exponential increase in the number of trials. Their main objective is to significantly reduce the number of experiments required while retaining sufficient information to analyse the main effects and, in some cases, certain critical interactions.

IV.2. Fractional factorial designs at two-level 2^{k-q}

Fractional factorial designs are an optimised experimental approach designed to minimise the number of trials while still allowing main effects to be estimated. These designs offer the possibility of conducting a study involving k factors with a significantly smaller number of experiments than the 2^k trials required by a full factorial design, while preserving an essential amount of statistical information. A fractional factoriel design has : $\boldsymbol{N = 2^{k-q}}$ trials.

$$\begin{cases} q = 1: N = 2^{k-1} trials \rightarrow 2^1 \ times \ fewer \ tests \ than \ the \ corresponding \ FFD; \\ q = 1: N = 2^{k-2} trials \ \rightarrow 2^2 \ times \ fewer \ tests \ than \ the \ corresponding \ FFD; \\ \qquad\qquad\qquad\qquad\qquad\qquad \ldots\ldots\ldots. \\ \qquad\qquad\qquad\qquad\qquad\qquad \ldots\ldots\ldots. \\ q : N = 2^{k-q} trials \ \rightarrow 2^q \ \ times \ fewer \ tests \ than \ the \ corresponding \ FFD. \end{cases}$$

with :

k : number of factors ;

2 : each factor takes on two levels (-1) and (+1);

q : means that the number of trials in the plan has been divided by 2^q.

Example IV. 1

A 25-2 fractional design studies 5 factors at 2 levels and is completed in only 8 trials compared to a full factorial design (FFP: 25 =32 trials). This means that the PFC has been divided by 22 = 4.d.

Remark

As q increases, the experimental load decreases, but at the expense of an increasing risk to the quality of the information obtained from the design. It is therefore necessary to assess the risks before starting the experiment and to minimise them by constructing the appropriate fractional design.

IV.3. Design of a fractional factorial experiment

As with full factorial designs, fractional designs are presented by mathematical models of polynomial form containing 2^k coefficients: a constant, main effects and interactions. If we carry out $N=2^{k-q}$ experiments, we obtain a system of N equations with M unknowns, with M > N .

Let be the matrix system : $Y = X.A$ $(\boldsymbol{Eq.\,IV-1})$

with : $\begin{cases} Y\ (N,1) \\ X\ (N,M) \\ A\ (M,1) \end{cases}$ et $M > N$

Mathematically, this system of equations is impossible to solve and in order to find its solution, we recommend adapting another model containing only N unknowns. So, how can this be done?

IV.3.1. Notion of contrasts and aliases

To reduce the number of unknowns M to N, it is advisable to group the coefficients of the initial model into new unknowns called contrasts (contrast theory). By definition, a contrast represents an apparent effect $\boldsymbol{L_i}$, of the reduced mathematical model which groups and combines two real effects (a main effect $\boldsymbol{a_i}$and an interaction $\boldsymbol{a_{ij}}$). These effects are said to be merged or aliased in the contrast. The use of contrast theory leads us to a new matrix system written in this form:

$$Y = X_S . L \qquad (\boldsymbol{Eq.\,IV-2})$$

$$\text{with : } \begin{cases} Y\ (N,1); \\ X_S\ (N,N): orthogonale\ matrix\ of\ Hadamard; \\ L\ (N,1): contrastes\ vector. \end{cases}$$

The transition from the system (Eq. IV-1) to the system (Eq. IV-2) generates a matrix calculation which makes it possible to determine the contrasts $\boldsymbol{L_i}$ of the reduced model as a function of the coefficients of the initial model (starting model). To do this, the effects matrix X(N, M) is decomposed into two sub-matrices: X_S(N, N) et X_I(N, M-N)

We have : $X = [X_S \vdots X_I]$ et $A = \begin{bmatrix} A_S \\ A_I \end{bmatrix}$

Equation (Eq. IV-1) can then be written as :

$$Y = X . A = [X_S \vdots X_I] \times \begin{bmatrix} A_S \\ A_I \end{bmatrix} = X_S . A_S + X_I . A_I \qquad (\boldsymbol{Eq.\,IV-3})$$

Combining (Eq. IV-2) and (Eq. IV-3), we obtainrons :

$$Y = X_S . L = X_S . A_S + X_I . A_I \qquad (\boldsymbol{Eq.\,IV-4})$$

The contrast vector $\boldsymbol{L}$ is obtained from the following equation: :

$$L = A_S + [X_S^T X_S]^{-1} X_S^T X_I . A_I \qquad (\boldsymbol{Eq.\,IV-5})$$

with : $[X_S^T X_S]^{-1} X_S^T X_I$ aliases matrix.

As mentioned above, a fractional design certainly saves on experimental costs, but it does have its drawbacks. Ambiguities can arise when interpreting the results, essentially due to the pairing of the coefficients in the initial model, which does not allow us to obtain the exact information or to extract the information we are looking for about the phenomenon under study. So how can these ambiguities be resolved and the mathematical model postulated in this type of plan interpreted properly?

IV.3.2. Notion of alias generator

To overcome this problem of interpretation, we recommend the use of the Box calculation, which uses an alias generator and a definition relation to quickly predict the pairs of

aliases of the coefficients of the initial model and the way in which they are aliased in the contrasts, without using the alias matrix.

By definition, an alias generator represents the highest order interaction of the initial model $\boldsymbol{x_{i\ldots k}}$ which, after a selective and successive sorting of (+1) and (-1) in the FFD's effects matrix, makes it possible to determine the set of aliases thanks to the definition relation which associates it with the $\boldsymbol{x_0}$ column (noted $\boldsymbol{I}$: element signifying +1), useful for calculating the average of the responses $\boldsymbol{a_0}$. The definition relation is written in the form: $\boldsymbol{I = x_{i\ldots k}}$.

Generally speaking, Box's calculation leads us to the following relationship:

$$L_i = a_i \pm \sum_{i}^{k} a_{i..k}$$

Remark :
The Box calculation is used to find columns with identical signs. It is only valid for two-level fractional factorial designs whose experimental points are located exactly at the vertices of the study domain.

IV.3.3. Interpretation assumptions

As we do not carry out all the experiments in the complete design, we cannot obtain the value of all the interactions. We have to create additional replacement information ourselves. This information must be realistic and compatible with the study being carried out. It is introduced in the form of hypotheses and needs to be verified before the study is concluded. These hypotheses are :

1. interactions of order 3 or higher are considered negligible;
2. if a contrast is zero, this may mean that :
- the aliased effects and interactions are all zero ;
- the aliased effects and interactions compensate each other;
3. if two contrasts are small, it is assumed that their interaction is also small;
4. if one contrast is weak and the other is strong, their interaction is assumed to be weak;
5. if two contrasts are strong, be wary of their interaction, which may also be strong;

6. if two interactions of order two are aliased in a contrast of significant value (in which case it is not known which interaction is responsible for the observed effect, or whether the two interactions each have an effect), it will be essential to carry out additional tests.

IV.4 Application example

Consider the matrix of the example used to study complete factorial design 2^3 (adsorption of amoxicillin by a hydroxyapatite).

Table IV.1. Experiment matrix of full factorial design at two factors 2^3.

Run N	x_0	x_1	x_2	x_3	x_{12}	x_{13}	x_{23}	x_{123}	$y(\%)$
1	+1	-1	-1	-1	+1	+1	+1	-1	55.89
2	+1	+1	-1	-1	-1	-1	+1	+1	56.93
3	+1	-1	+1	-1	-1	+1	-1	+1	64.70
4	+1	+1	+1	-1	+1	-1	-1	-1	61.23
5	+1	-1	-1	+1	+1	-1	-1	+1	85.95
6	+1	+1	-1	+1	-1	+1	-1	-1	89.95
7	+1	-1	+1	+1	-1	-1	+1	-1	88.50
8	+1	+1	+1	+1	+1	+1	+1	+1	96.55

Table IV.2. Coefficient values for the AMX adsorption model.

constant	Effets linéaires			Effets d'interactions			
a_0	a_1	a_2	a_3	a_{12}	a_{13}	a_{23}	a_{123}
74.96	1.20	2.78	15.28	-0.06	1.81	-0.49	1.07

To split this complete factorial design into two parts in order to obtain a fractional 2^{3-1} design, it is necessary to :

1. sort the rows of this design according to the values in the highest interaction column (Box calculation). In this case, the alias generator is the triple interaction 123 ;

The sorting is done in relation to the (+1) and (-1) in order to obtain a part of the upper half-plane containing only (+1) in the number of 4 rows in column 123 and a part of the lower half-plane containing only (-1) in the number of 4 (Table III.2).

Table IV.3. Matrix of X (N, M) effects sorted by FFD

N	x_0	x_1	x_2	x_3	x_{12}	x_{13}	x_{23}	x_{123}	y	
2	+1	+1	-1	-1	-1	-1	+1	+1	55.89	Plan upp
3	+1	-1	+1	-1	-1	+1	-1	+1	56.93	
5	+1	-1	-1	+1	+1	-1	-1	+1	64.70	
8	+1	+1	+1	+1	+1	+1	+1	+1	61.23	
1	+1	-1	-1	-1	+1	+1	+1	-1	85.95	Plan low
4	+1	+1	+1	-1	+1	-1	-1	-1	89.95	
6	+1	+1	-1	+1	-1	+1	-1	-1	88.50	
7	+1	-1	+1	+1	-1	-1	+1	-1	96.55	

2. Choose an upper or lower half-plane and carry out the study in relation to this part (carry out the experiments based on the chosen part). In the following, we will use the upper half-plane in our calculations;

3. Determine the aliased columns of the new effects matrix which have a sequence of signs (+1) and (-1) which are identical 2 by 2 (Example: column 1 and column 123, which is equivalent to writing 1=123);

4. Determine the contrasts (apparent effects resulting from the fact that the effects are merged 2 by 2).

- **Upper half-plane:**

Aliases generator is: column 123.

Definition relationship: I=123

According to Box, its calculation is based on the following relationships:

- multiplying a column by itself gives a sign column (+); I: $1 \times 1 = 1^2 = I$;

- multiplying a column by a column of signs (+); I, gives back the initial column: $1 \times I = I \times 1 = 1$.

Multiply the 2 members of the definition relation I = 123 by 1; we obtain $1 \times I = 1 \times 123 = 1 \times 1 \times 23 = 1^2 \times 23$ and simplifying according to the rules above (Box algebra): 1 = 23; which means that the main effect 1 is therefore aliased with the interaction 23.

Doing the same with all the factors, we obtain :

2 = 13; the main effect 2 is aliased with the interaction 13;

3 = 12; main effect 3 is aliased with interaction 12.

Once the aliases have been found, we determine the L_i contrasts:

$$L_1 = L_{23} = a_1 + a_{23}$$
$$L_2 = L_{13} = a_2 + a_{13}$$
$$L_3 = L_{12} = a_3 + a_{12}$$
$$L_0 = L_{123} = a_0 + a_{123}$$

The system obtained after reducing the number of unknowns M is as follows:

Table IV.4. Matrix of X (N, N) effects

N		x_0	x_1	x_2	x_3	**y**
2		+1	+1	-1	-1	55.89
3		+1	-1	+1	-1	56.93
5		+1	-1	-1	+1	64.70
8		+1	+1	+1	+1	61.23

Plan upp

Table IV.5. Apparent coefficient values for the AMX adsorption model.

Apparent effects	L_0	L_1	L_2	L_3
vales	76.03	0.71	4.59	15.22

Remark :

In the lower half-plane, column 123 contains only (-1) values, whereas column I contains only (+1) values, giving us the defining relationship: I = - 123. From here, we can find all the aliases by posing: $1 \times I = -1 \times 123 = -1^2 \times 23$, or 1 = -23 for example.

Once the contrast values have been determined, a statistical analysis is run to validate the postulated mathematical model before it is interpreted.

Chapter V
Response surfaces methodology

Chapter V : Response surfaces methodology

V.1. Introduction

The first experimental designs published in the literature were designed to estimate the effects of factors. However, from an industrial point of view, not all the problems encountered consist solely of estimating and then comparing the effects of factors. Many studies involve finding an optimum, if any, in a field of study known as the experimental field. For this purpose, the methodology of experimental design for response surfaces was created in order to determine quantitatively the variations in response with respect to the factors of significant influence, identified if necessary during an initial screening study.

V.2. Response surfaces methodology

The study of response surfaces is associated with the use of second-degree polynomials, which can be written in the following general form :

$$\hat{y} = b_0 + \sum_{j=1}^{k} b_j x_j + \sum_{j=1}^{k} \sum_{u \neq j}^{k} b_{ju}\, x_j x_u + \sum_{j=1}^{k} b_{jj}\ x_j^2 \qquad (\boldsymbol{Eq}.\mathbf{V}-\mathbf{1})$$

The plans corresponding to this type of model must :

- allow each of the factors to be studied on at least three levels ;
- include a reasonable number of trials ;
- satisfy the optimality criteria, given that the matrix X can no longer be orthogonal.

There are several types of second-degree design, the main ones being :

V.2.1. Central composite design

Centred composite design consists of carrying out the tests of a factorial design supplemented by experiments at the centre of the field of study and star tests (Figure V.1). The total number of trials N to be carried out is the sum of :

- $\boldsymbol{n_f}$ full factorial design trials ;
- $\boldsymbol{n_\alpha = 2k}$ star tests on axes at a distance α from the centre of the domain ;
- $\boldsymbol{n_0}$ trials at the centre of the domain.

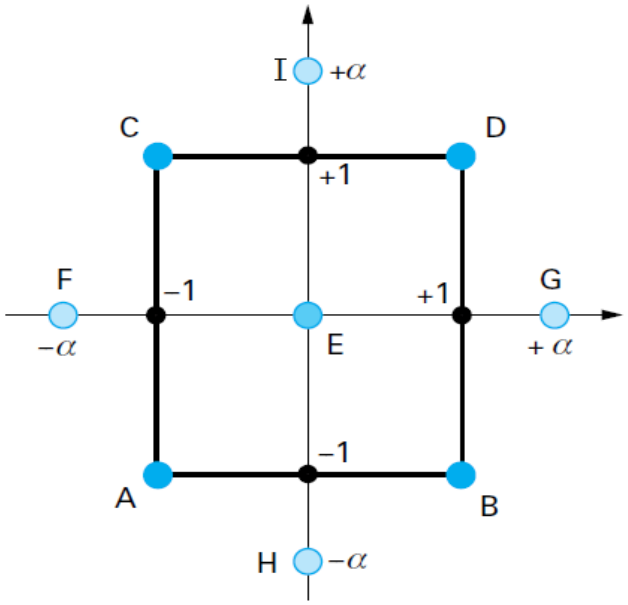


Figure V.1. Representative diagram of a centred composite design for two factors

The figure above illustrates the study domain of a two-factor composite design. Points A, B, C and D are the experimental points of a full 2^2 factorial design. Point E is the central point. This point can be repeated one or more times. Points F, G, H and I are the axial points located at a particular distance α from the centre of the study domain (α plays an important role in the quality of the model and should not be chosen at random). These last four points form what is known as the star plane.

Furthermore, the classical mathematical model derived from the general model (Eq. V-1) and representing the two-factor composite design is written as :

$$\hat{y} = a_0 + a_1x_1 + a_2x_2 + a_{12}x_1x_2 + a_{11}x_1^2 + a_{22}x_2^2$$

V.2.1.1. Properties of composite designs

- **Orthogonality of the calculation matrix**

  For a two-factor composite design, 9 experiments are required to determine the values of the 6 coefficients of the equation of the postulated second-degree model. Thus, the effects matrix is as follows:

$$X = \begin{vmatrix} + & - & - & + & + & + \\ + & + & - & - & + & + \\ + & - & + & + & + & + \\ + & + & + & - & + & + \\ + & 0 & 0 & 0 & 0 & 0 \\ + & -\alpha & 0 & 0 & +\alpha^2 & 0 \\ + & +\alpha & 0 & 0 & +\alpha^2 & 0 \\ + & 0 & -\alpha & 0 & 0 & +\alpha^2 \\ + & 0 & +\alpha & 0 & 0 & +\alpha^2 \end{vmatrix}$$

As the matrix X is not orthogonal, the matrix $[X^T X]^{-1}$ is no longer diagonal. The variances of the model coefficients are obtained by multiplying the elements $\boldsymbol{C_{jj}}$ of the diagonal of this dispersion matrix by the reproducibility variance :

$$S_{bj}^2 = C_{jj}.S_{rep}^2 \qquad (\boldsymbol{Eq.V-2})$$

The solution vector is calculated in the same way as for the factorial plane using the following equation :

$$A = [X^T X]^{-1} X^T Y$$

- **Optimality criteria**
  - ✓ **Rotational isovariance criterion:** for the composite design to satisfy this criterion, the star points must be placed at a distance α equal to the fourth root of the number of points in the factorial design: $\alpha = n_f^{1/4}$;
  - ✓ **Uniform precision criterion:** obtained by increasing the number of central points. Its purpose is to ensure the same predicted response accuracy throughout the domain;
  - ✓ **Orthogonality criterion:** the orthogonality criterion is met if the matrix $[X^T X]^{-1}$ is diagonal. In the case of centred composite planes, this criterion is met if α chosen satisfies the following condition :
  - ✓ $\alpha = \left(\frac{n_f\left(\sqrt{N}-\sqrt{n_f}\right)^2}{4}\right)^{\frac{1}{4}}$

The values of the parameter α and the number of points at the centre n_0 depend on the number of factors k in the basic factorial design and the optimality criterion fulfilled by the design (Table V.1).

Table V.1. Values of α and n_0 according to the properties equired for CCD.

k		2	3	4	5	$2^{(5-1)}$	6	$2^{(6-1)}$
n_f	2^k (ou 2^{k-p})	4	8	16	32	16	64	32
n_α		4	6	8	10	10	12	12
	Rotational isovariance	≥1	≥1	≥1	≥1	≥1	≥1	≥1
n_0	Uniform precision	5	6	7	10	6	15	9
	ortogonality	8	12	12	17	10	24	15
α		1,41	1,68	2	2,38	2	2,83	2,38

Remark :

The major disadvantage of the composite design strategy is that it requires more experimental points than coefficients to be determined.

V.2.2. Box-Behnken design

The Box-Behnken design is economical and requires only 3 levels per factor, unlike the composite design, which requires 5 levels. The Box-Behnken design is a second-degree design which does not rely on the full factorial design stage. It is used when the researcher is sure of knowing all the factors influencing the phenomenon being studied. He starts with three factors, but can use four, five or more.

The Box-Behnken plan has the property of sequentiality with respect to the factors, i.e. it is possible to study k factors while reserving the possibility of adding others, while retaining the results of tests already carried out.

For a number of three factors, the Box-Behnken design is constructed as follows (Figure V.2):

- ✓ write a 22 design on two factors and set the remaining k-2 variables to level 0 (in this case k-2= 3-2=1) ;

- ✓ repeat the operation for all possible combinations of factors (in this case, repeat twice);
- ✓ add points to the centre.

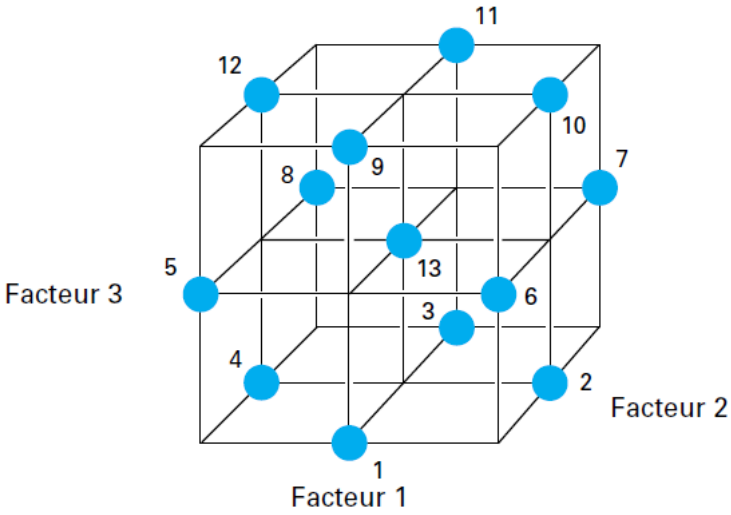


Figure V.2. Box-Behnken design for three factors

Figure V.2 illustrates the study domain of a three-factor Box-Behnken design. The experimental points (points 1 to 12) are all placed at equal distances from the centre of the study area: the middle of the edges of the cube. Point 13, on the other hand, represents the test at the centre that Messrs Box and Behnken recommend repeating three times. This design therefore comprises twelve trials, to which we can add one (or more) central point(s). Table IV.2 shows these twelve trials with 3 central points. The construction of the matrix of the Box-Behnken plane gives it the particularity of being isovariant by rotation (spherical symmetry).

As with the composite plane, the classical mathematical model derived from the general model (Eq. V-1) and representing the three-factor Box-Behnken plane is written :

$$\hat{y} = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{23}x_2x_3 + a_{11}x_1^2 + a_{22}\,x_2^2 + a_{33}\,x_3^2$$

Combining the study area schematised in Figure V.2 (Table V.2) and the representative mathematical model for a number of factors equal to three, it emerges that a Box-Behnken design requires 13 to 15 trials to determine 10 coefficients. It is therefore necessary to solve a system of 13 to 15 equations with 10 unknowns.

Table V.2. Box-Behnken matrix for 3 factors.

N	x_1	x_2	x_3
1	-1	-1	0
2	+1	-1	0
3	-1	+1	0
4	+1	+1	0
5	-1	0	-1
6	+1	0	-1
7	-1	0	+1
8	+1	0	+1
9	0	-1	-1
10	0	+1	-1
11	0	-1	+1
12	0	+1	+1
13	0	0	0
14	0	0	0
15	0	0	0

Remark :

To determine the total number of trials N to be carried out, it is preferable to first adopt the second-degree mathematical model corresponding to the number of factors studied, and then count the number of interactions in this model.

As the matrix X of the Box-Behnken design is not orthogonal, the matrix $[X^T X]^{-1}$ is no longer diagonal. This amounts to doing the same calculation as the composite plane to determine the coefficients of the postulated model.

V.3. Notions of response surfaces and isoresponse curves

Once the second degree mathematical model has been established, it is used to find the optimum combination of operating parameters leading to the best response y. The optimum can be found either by mathematically solving the equation representing the phenomenon under study, or by graphically solving it using response surfaces and isoresponse curves.

V.3.1. Response surfaces

The second degree model obtained after experimentation can be used to plot the relationship $\hat{y} = f(x_1, x_2, \ldots . x_k)$ in the space of the variables $x_1, x_2, \ldots . x_k$. This relationship is therefore illustrated by a surface called the response surfaces, which is nothing other than a set of graphs represented in a three-dimensional space (Figure IV.3); two dimensions for the factors and one for the response. Thus, the horizontal plane of the figure represents the range of variation of two factors and the vertical axis represents the variation of the response based on the postulated model. It should be noted that the shape of the response surfaces (paraboloid or hyperboloid) can tell us whether there is a maximum or a minimum, or both. However, their curvatures are due to the presence of square terms in the model.

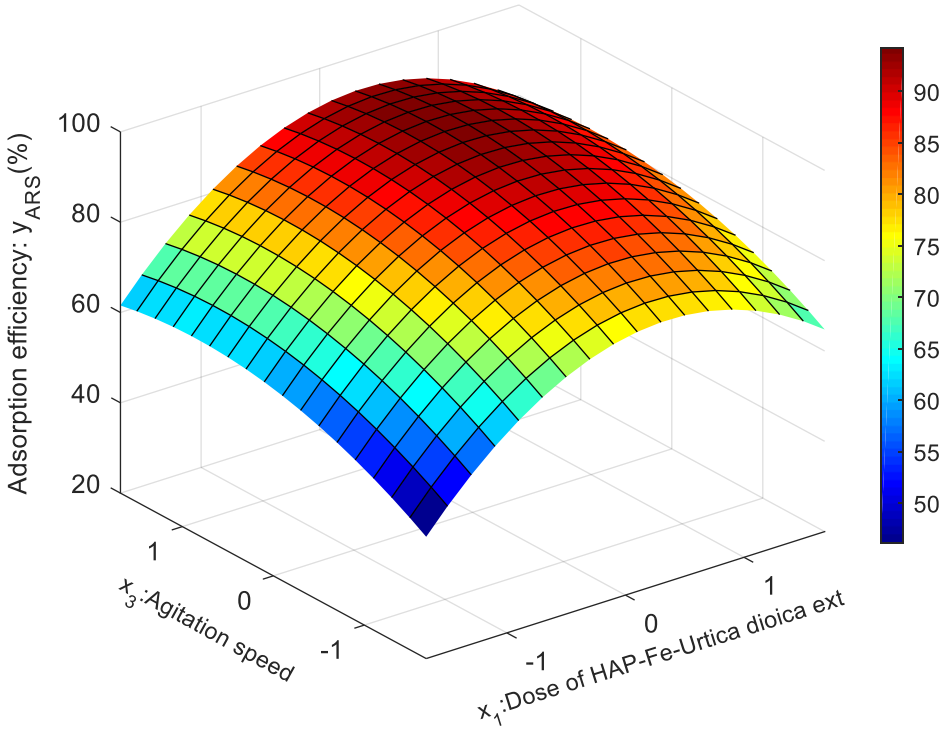


Figure V.3. Response surface curve.

V.3.2. Isoresponse curves

The isoresponse curves interpret the projection of all the points making up the response surface in the horizontal plane of the variables $x_1, x_2, \ldots . x_k$ (Figure V.4). They can be interpreted as contour lines onto which the value of the response is projected.

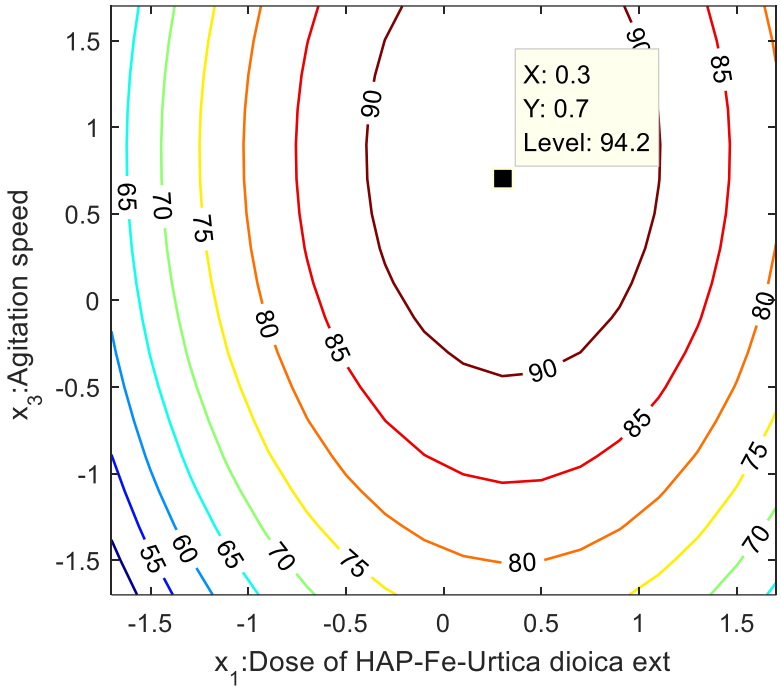


Figure V.4. Isoresponse curve.

V.4 Application example

V.4.1. Problem to be solved

A researcher wants to use shoe sole waste to recover an organic pollutant: bezacryl yellow (JB) in a closed stirred reactor. His main interest is in obtaining the optimum conditions for pH, JB concentration and shoe sole waste concentration (in the form of a suspension), which will give a better adsorption rate for bezacryl yellow. The researcher used a centred composite design for this study. The ranges of variation for each of the parameters studied are : [JB]∈[10-50]mg/L , [susp]∈[0.1-1] g/250mL) and pH∈[2-10].

V.4.2. Problem formulation

- ✓ objective of the study: optimisation of the operating parameters for a better adsorption yield of bezacryl yellow by the waste shoe soles ;
- ✓ define the factors and the response :

Response y: rate of adsorption of bezacryl yellow by the adsorption process, calculated as follows:

$$y\,(\%) = \frac{[JB]_0 - [JB]_t}{[JB]_0} \times 100$$

With :

$[JB]_0$: initial concentration of bezacryl yellow ;

$[JB]_t$: concentration of the bezacryl yellow solution at time t.

Factors and ranges of variation :

Factor 1 ⇒ z_1: concentration of bezacryl yellow, [JB]∈[10-50]mg/L ;

Factor 2 ⇒ z_2: concentration of shoe sole suspension; [susp]∈[0.1-1] g/250mL ;

Factor 3 ⇒ z_3: pH of the solution; pH∈[2-10]

To conduct a study based on a three-factor centred composite design with uniform precision, we need $N = n_f + n_\alpha + n_0 = 8 + 6 + 6 = 20$ experiments to be performed. The mathematical model associated with this design is a model with M =10 coefficients which is written in the following form:

$$\hat{y} = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{23}x_2x_3 + a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2$$

The levels of the factors and their correspondence in coded variables are listed in Table V.3.

Table V.3. Values of operating parameters at different levels.

Real variables	Coded variables	Low level		Central	High level	
		-1.68	-1	0	+1	1.68
$z_1 : [JB]_0$ (mg/L)	x_1	10	18.1	30	41.9	50
$z_2 : [Susp]_0$ (g/250mL)	x_2	0.1	0.29	0.55	0.81	1
$z_3 : pH$	x_3	2	3.62	6	8.38	10

The values -1.68 and +1.68 are respectively, the low and high level for the axial α points for a number of factors k=3.

V.4.3. Planning and experimentation

- Construction of the design of experiments: construct the matrix of experiments in real and coded units (Table V.4) in order to use it as a start-up tool to launch and facilitate the implementation of the experiment.

Table V.4. Matrix of experiments for a centred composite design.

N	z_1	z_2	z_3	x_0	x_1	x_2	x_3	y **(%)**
1	18,1	0,29	3,62	+1	-1	-1	-1	93,92
2	18,1	0,29	8,38	+1	+1	-1	-1	82,45
3	18,1	0,81	3,62	+1	-1	+1	-1	70,11
4	18,1	0,81	8,38	+1	+1	+1	-1	97,70
5	41,9	0,29	3,62	+1	-1	-1	+1	85,52
6	41,9	0,29	8,38	+1	+1	-1	+1	87,85
7	41,9	0,81	3,62	+1	-1	+1	+1	87,40
8	41,9	0,81	8,38	+1	+1	+1	+1	86,42
9	30	0,55	6	+1	0	0	0	88,26
10	30	0,55	6	+1	0	0	0	91,63
11	30	0,55	6	+1	0	0	0	89,30
12	30	0,55	6	+1	0	0	0	87,40
13	30	0,55	6	+1	0	0	0	90,50
14	30	0,55	6	+1	0	0	0	90,00
15	10	0,55	6	+1	-1,68	0	0	81,20
16	50	0,55	6	+1	+1,68	0	0	85,82
17	30	0,1	6	+1	0	-1,68	0	94,30
18	30	1	6	+1	0	+1,68	0	90,00
19	30	0,55	2	+1	0	0	-1,68	23,43
20	30	0,55	10	+1	0	0	+1,68	80,36

V.4.4. Analysis of the results

- **Calculation of model coefficients**

As mentioned previously, the coefficients of the regression equation are estimated by calculating the following matrix product: $B = [X^T X]^{-1} X^T Y$. The results of this calculation are shown in Table V.5.

Table V.5. Values of the coefficients of the second-degree model.

Constant		Linear effects					
a_0		a_1		a_2		a_3	
89.13		1.85		-1.12		7.23	
Interaction effects							
a_{12}	a_{13}	a_{23}		a_{11}	a_{22}		a_{33}
4.47	-1.85	1.13		0.30	3.35		-10.91

- **Statistical analysis of the regression equation**
  - **Checking the significance of the coefficients**

The value of the reproducibility variance is S^2_{rep}= 2.35. However, to determine the variance of the model coefficients, we simply multiply the elements C_{jj} of the diagonal of the dispersion matrix $[X^T X]^{-1}$ (EXCEL) by the reproducibility variance : $S^2_{bj} = C_{jj} S^2_{rep}$. The tj values obtained for each effect are shown in Table V.6.

Table V.6. Values of tj

Constant		Linear effects					
t_0		t_1		t_2		t_3	
227.61		10.73		6.51		41.90	
Interaction effects							
t_{12}	t_{13}	t_{23}		t_{11}	t_{22}		t_{33}
15.18	6.27	3.83		1.80	20.46		66.56

For the significance level α = 0.05 and the number of degrees of freedom f = (n_0 - 1) = 5, the tabulated value of Student's t_α (f) is equal to 2.57. As the value of t_{11} is less than tα (f), the coefficient a_{11} is therefore insignificant and therefore eliminated from the regression equation.

- **Validation test of the regression equation**

Search for bias

The calculated value of the residual variance for N = 20 and ℓ = 9 is equal to Sres2 = 142.26, which gives: F = 1.59. The tabulated value of the Fisher Snedecor test for significance level α = 0.05 and numbers of degrees of freedom (N - ℓ) = 11 and (n_0 - 1) = 5 varies between 4.74 and 4.68. As this interval is greater than the calculated value, the model is considered unbiased.

Significance test for regression

For the significance level α = 0.05 and the numbers of degrees of freedom (ℓ -1) = 8 and (N-ℓ) = 11, the value of the tabulated Fisher test is 2.95. Since the calculated value of F is greater than the tabulated value, the regression equation is adequate and the model is validated at 95%. The equation used for the model is therefore written as :

$$\hat{y} = 89{,}13 + 1{,}85\,x_1 - 1{,}12x_2 + 7{,}23x_3 + 4{,}74x_1x_2 - 1{,}85x_1x_3 + 1{,}13x_2x_3 + 3{,}35\,x_2^2 - 10{,}91\,x_3^2$$

Model analysis indicates that the most influential parameters are initial dye concentration and solution pH with positive effects of 1.85 and 7.23 respectively. Suspension concentration had a negative effect of -1.12.

- **Analysis of residuals**

The quality of the second-order model can be assessed by analysing the residuals (Figure V.5).

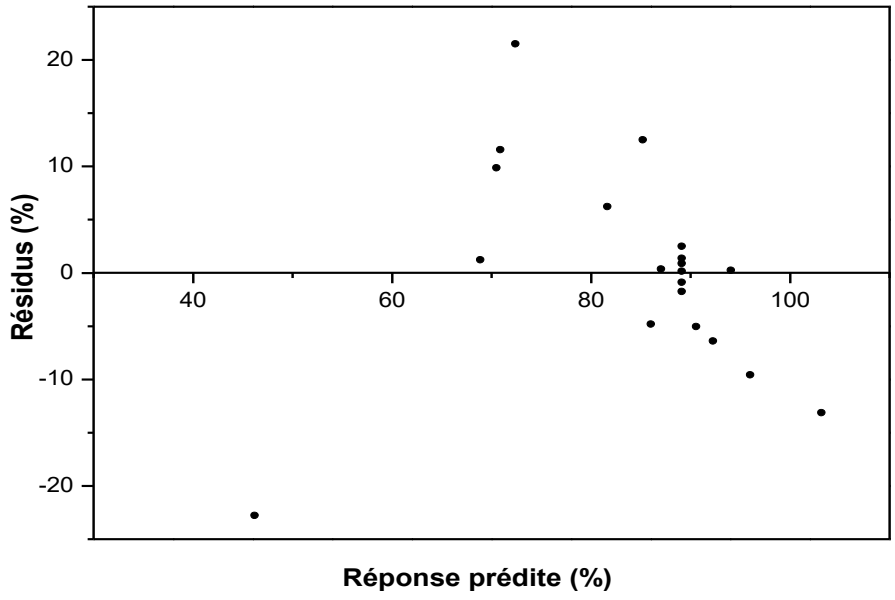


Figure V.5. Residuals diagram

The diagram in Figure IV.5 shows no relationship between the predicted values and the residuals, as the points appear to be arranged randomly. This result reflects the absence of information in the residuals, so all the information is explained by the regression model used. In conclusion, the second-order model obtained simulates perfectly the JB adsorption process on the sole waste; it will therefore be used to optimise the operating parameters.

V.4.5. Optimisation

After the validation stage, which enabled us to arrive at a second degree model translating the effect of the variation of the three operating parameters studied on the rate of adsorption of JB by the shoe sole waste, the search for the optimum values of these parameters can then be carried out either by :

- **Mathematical resolution**

It should be remembered that the modelling of the process studied led to the following equation:

$$\hat{y} = 89{,}13 + 1{,}85\,x_1 - 1{,}12x_2 + 7{,}23x_3 + 4{,}74x_1x_2 - 1{,}85x_1x_3 + 1{,}13x_2x_3 + 3{,}35\,x_2^2 - 10{,}91\,x_3^2$$

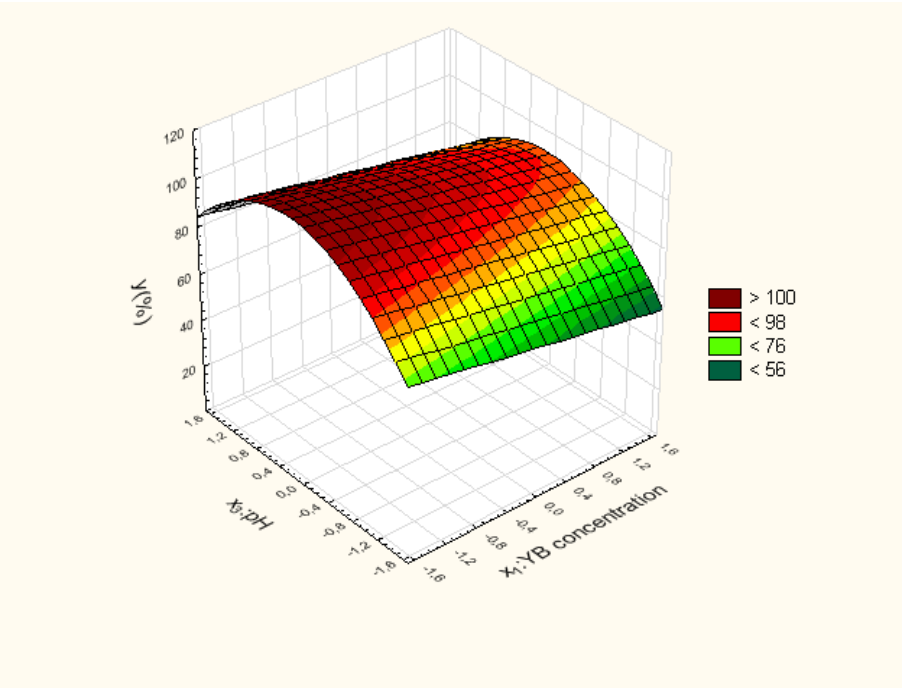
This predictive mathematical equation is used to calculate the optimum values of the operating parameters leading to the optimum response in the field of study. Simply derive the function with respect to each of the variables $\boldsymbol{x_1}$, $\boldsymbol{x_2}$ and , $\boldsymbol{x_3}$. The system of equations obtained after derivation is represented as follows :

$$\begin{cases} \dfrac{\partial \hat{y}}{\partial x_1} = 1{,}85 + 4{,}74x_2 - 1{,}85x_3 = 0 \\ \dfrac{\partial \hat{y}}{\partial x_2} = -1{,}12 + 4{,}74x_1 + 1{,}13x_3 + 6{,}70\,x_2 = 0 \\ \dfrac{\partial \hat{y}}{\partial x_3} = 7{,}23 - 1{,}85x_1 + 1{,}13x_2 - 21{,}82\,x_3 = 0 \end{cases}$$

This system cannot be solved analytically, so a numerical solution is required.

- **Graphical resolution**

The response surface curves plotted using the STASTICA software show a maximum, which indicates the existence of a maximum efficiency.



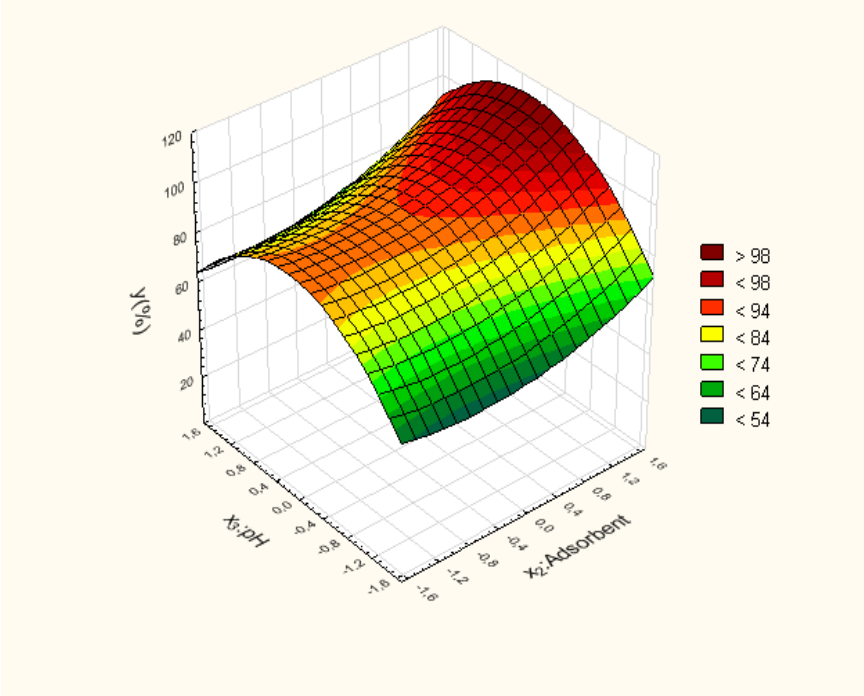


Figure V.6. Response surface curves.

According to these curves, a JB adsorption rate of around 98% (red zone) is obtained when working with a JB concentration equal to 30 mg/L, a shoe sole suspension concentration equal to 0.1 g/250 mL and a pH equal to 6. This combination of three parameters represents the optimum conditions sought.

Chapter VI
Mixture Design

Chapter VI : Mixture Design

VI.1. Introduction

Unlike conventional experimental designs (complete and fractional factorial designs, response surface designs, etc.) which are based on the study of the influence of independent factors on the measured response, mixture designs are special designs, adapted to dependent factors.

VI.2. Fundamental constraint of mixtures

If we denote by x_i the content of constituent i in any mixture, the sum of the contents of all the constituents in the mixture satisfies the relation/

$$\sum_{i=1}^{N} x_i = 100\% \qquad \text{or} \qquad \sum_{i=1}^{N} x_i = 1 \qquad (\boldsymbol{Eq.VI-1})$$

In this way, the content of each component remains between 0 and 100%. When the content of one of the constituents is increased, the other contents automatically decrease so that the sum of all the contents remains equal to 100%.

This relationship, known as the fundamental constraint of mixtures, has important consequences for the geometric and matrix representations of these planes. It also has repercussions on the modelling of results.

VI.3. Unconstrained mixture designs

In this type of design, the mixtures studied are made up of products whose contents can vary from 0 to 1.

VI.3.1. Geometric representation

As with experimental designs, Cartesian axes can be used to graphically present the composition of the mixture studied. The first axis is assigned to the first component (factor 1: x_1). The second axis, orthogonal to the first, is assigned to the second component (factor 2: x_2) and so on. Figure V.1 illustrates the Cartesian representation of a mixture of three components A, B and C. Figure VI.1.(a) shows the location of a point in a three-component mixture, while Figure VI.1.(b) reveals the set of three-component mixtures in the Cartesian space of the three components A, B and C.

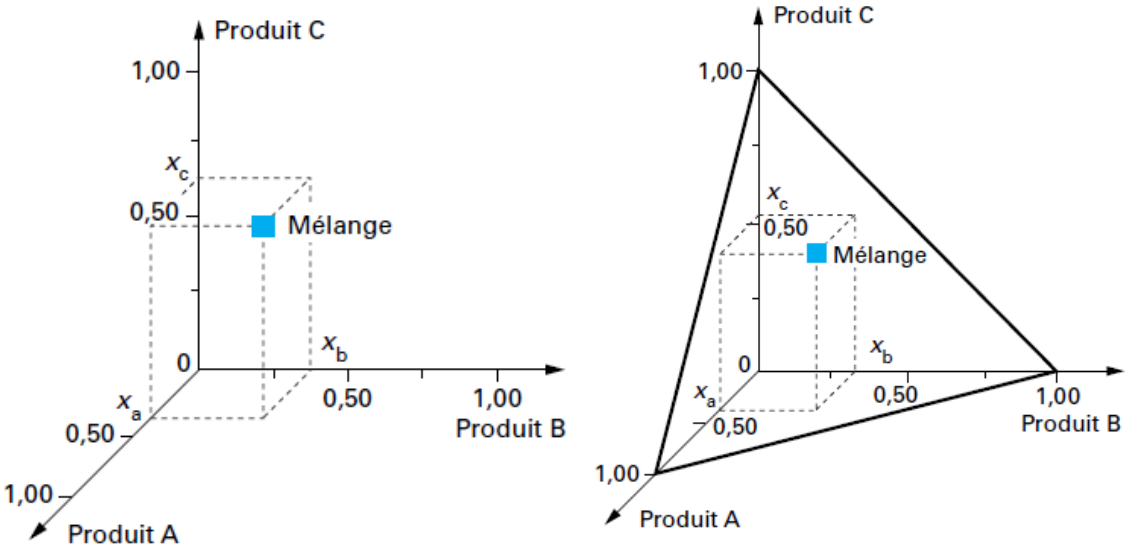


Figure VI.1. Representation of three-component mixtures in Cartesian space.

It should be noted that the geometric representation of a mixture must also take into account the fundamental constraint of mixtures (Eq. VI-1,). This causes a dimension to disappear in the x_i space. Thus, to define a study domain for a mixture composed of two constituents A and B (binary), we use a straight line segment (Figure VI.2.(a)) and for a mixture with three constituents A, B and C (ternary), we adopt the representation of an equilateral triangle whose vertices are the three points with abscissae 1 (Figure VI.2.(b)). In fact, the pure constituents are represented by the ends of the right-hand segment for the binary and the vertices of the equilateral triangle in the case of the ternary. For example, a binary blend with 50% A is located in the middle of the segment delimited by components A and B. Point M with coordinates x_a, x_b et x_c on the other hand, corresponds to a tertiary blend.

Reamark :

The vertices of the equilateral triangle correspond to pure components;

The sides of the triangle correspond to binary mixtures;

The inside of the triangle corresponds to all ternary mixtures.

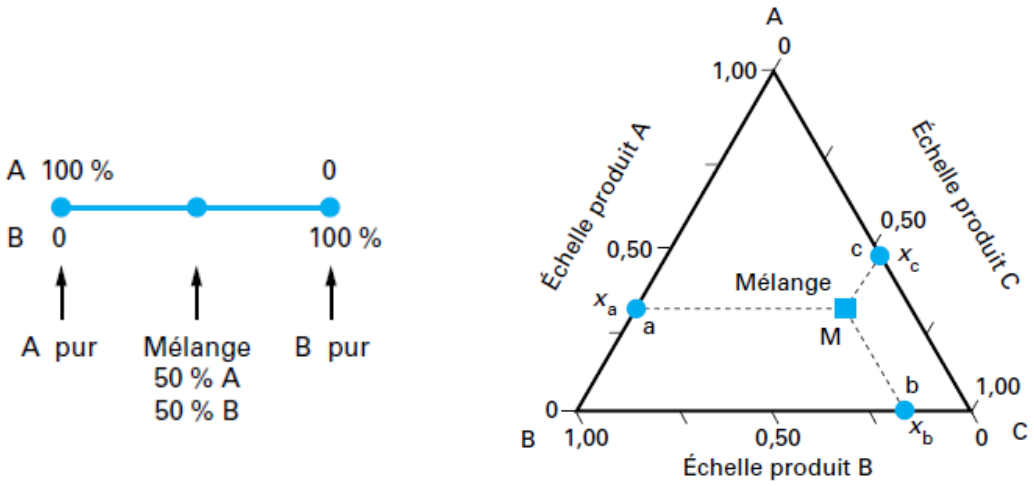


Figure VI.2. Study area for compound mixtures: (a): two constituents and (b): three constituents.

Beyond four constituents, the geometric representation becomes difficult, which necessitates the use of a table (matrix of tests) to group all the mixtures studied, as for classic experimental designs.

III.3.2. Location of experimental points

For a three-component mixture and when there are no constraints, the experimental points are distributed throughout the study area (Figure VI.3).

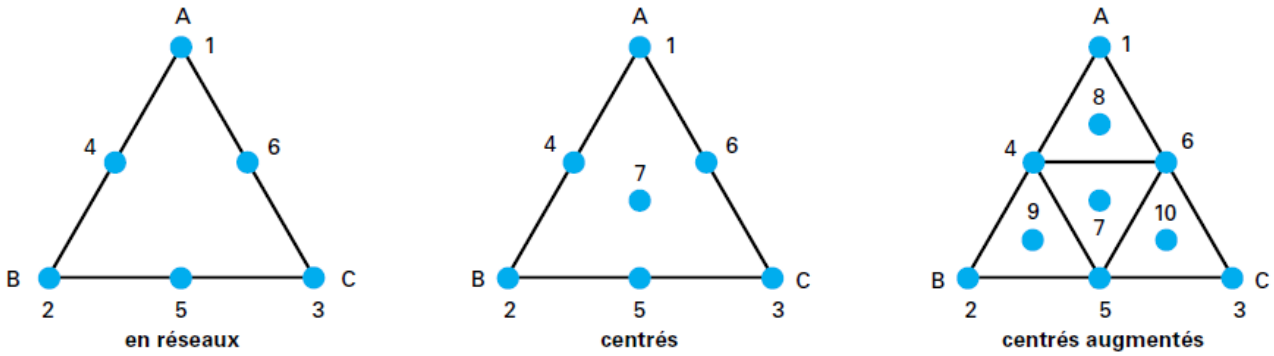


Figure III.3: Network mixing plane (left), centred mixing plane (middle), augmented centred mixing plane (right).

Depending on the arrangement of these points, several types of mixing scheme can be distinguished:

- Simplex lattice designs: these include the three pure components and half-and-half mixtures;

- centred designs (Simplex-Centroid Designs): these are distinguished from lattice designs by the systematic presence of a central point made up of 1/3 A, 1/3 B and 1/3 C ;
- Augmented Simplex-Centroid Designs: these are centred mixing designs to which the centres of gravity of the unit simplexes are added.

VI.4. Mixture designs with constraints

In this type of design, the mixtures are characterised by numerous constraints (low, high or relational constraints) which may affect the choice of constituent proportions. Depending on these constraints, the planning of the study is modified and must be adapted to each case.

VI.4.1. Prohibited low levels

This case reflects the fact that the composition of one or more constituents may not be lower than a given value: this is the lower limit noted L_i. This condition reduces the field of study while retaining its original shape (for a three-component mixture, we have an equilateral triangle: Figure VI.4.(a)).

VI.4.2. High prohibited contents

In this case, the composition of one or more constituents may not exceed a given value: this is the upper limit noted U_i. In this case, the field of study is reduced, but the shape of the field is completely modified (for a three-component mixture, it is no longer a triangle but a polygon: Figure VI.4.(b)).

III.4.3 Prohibited high and low contents

This constraint reveals the possibility that any composition of one or more constituents may be subject to both lower and higher constraints. The introduction of dual constraints generally has the effect of modifying the shape of the experimental domain. It can be an equilateral triangle, a trapezoid, a pentagon (Figure VI.4.(c)).

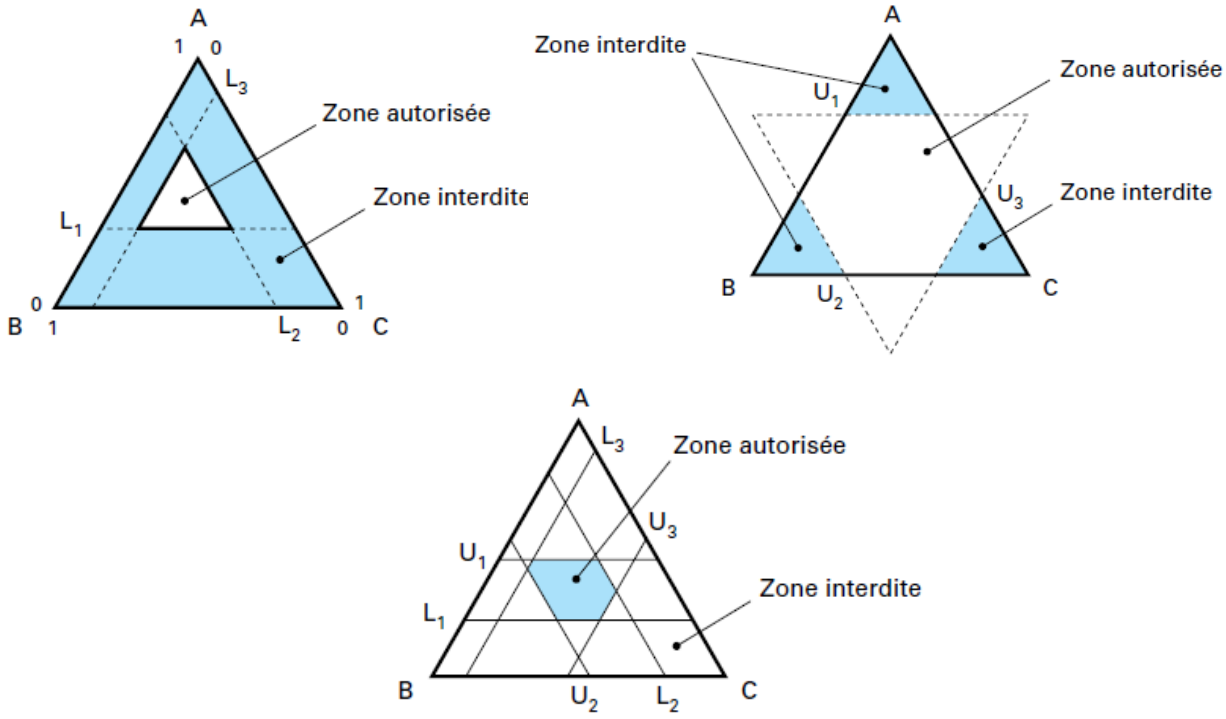


Figure VI.4. Representation of the study domains of mixtures with constraints.

VI.4.4 Relational constraints

In addition to the high and low forbidden content constraints, mixtures present other relational constraints linked either to :

- maintaining a constant ratio between the proportions of two constituents ;
- respecting an addition relationship between the proportions of two or more constituents.

These new constraints lead to new restrictions on the field of study and modify the location of the experimental points.

VI.4.5. Location of experimental points

The general approach adopted for selecting the experimental points of a domain of any shape follows from Scheffé's proposal for regular domains; the informative points are located at the vertices of the domain, at the midpoints of the edges, at the centres of the faces or hyperfaces and at the centre of gravity of the volume or hypervolume.

VI.5. Mathematical models associated with mixing designs

Like all classical experimental designs, to translate the variations of the property yˆ as a function of the composition of the mixture, the experimenters choose classical polynomial models which can be written in a simplified form called the canonical form and which results from the existence of the equality constraint on the fractions of constituents in the mixture.

VI.5.1. First degree model

For the study of a three-component mixture, the postulated first-degree mathematical model is obtained from a classical polynomial of order one, i.e. :

$$\hat{y} = a_0 + a_1x_1 + a_2x_2 + a_3x_3 \qquad (\boldsymbol{Eq.VI-2})$$

to which we apply the fundamental constraint of mixtures :

$$x_1 + x_2 + x_3 = 1$$

with :

$\hat{y}$: response calculated from the model at the chosen composition point ;

The relationship (Eq.III-2) can then be written as :

$$\hat{y} = a_0(x_1 + x_2 + x_3) + a_1x_1 + a_2x_2 + a_3x_3 \qquad (\boldsymbol{Eq.VI-3})$$

By grouping the coefficients :

$$\hat{y} = (a_0 + a_1)x_1 + (a_0 + a_2)x_2 + (a_0 + a_3)x_3$$

the model then takes the following form :

$$\hat{y} = b_1x_1 + b_2x_2 + b_3x_3 \qquad (\boldsymbol{Eq.VI-4})$$

with :

$\begin{cases} b_1 = a_0 + a_1 \\ b_2 = a_0 + a_2 \\ b_3 = a_0 + a_3 \end{cases}$: coefficients linéaires du modèle.

VI.5.2. Second degree model

The second-degree model with interactions (synergistic) is written in the form :

$$\hat{y} = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{23}x_2x_3 + a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 \qquad (\boldsymbol{Eq.\,VI-5})$$

becomes , after application of the fundamental constraint for mixtures :

$$\hat{y} = b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 \qquad (\boldsymbol{Eq.\,VI-6})$$

with :

$$\begin{cases} x_1 = (1 - x_2 - x_3) \\ x_1^2 = x_1(1 - x_2 - x_3) \\ x_1^2 = x_1 - x_1x_2 - x_1x_3) \end{cases}$$

and

$$\begin{cases} b_1 = a_0 + a_1 + a_{11} \\ b_2 = a_0 + a_2 + a_{22} \\ b_3 = a_0 + a_3 + a_{33} \\ b_{12} = a_{12} - a_{11} - a_{22} \\ b_{13} = a_{13} - a_{11} - a_{33} \\ b_{23} = a_{23} - a_{22} - a_{33} \end{cases}$$

The fundamental constraint of the mixtures makes the constant disappear and the second degree terms are reduced to the rectangular terms (double and triple interactions).

The coefficients of the postulated model are calculated from the regression relation applied in the case of classical experimental designs.

References

[1] J. Goupy, Introduction aux plans d'expériences, Edition DUNOD, Paris, 2001.

[2] G. Sado, M. C. Sado, Les plans d'expériences : De l'expérience à l'assurance qualité, Edition AFNOR TECHNIQUE, Paris, 1996.

[3] Z. R. Lazic, Design of Experiments in Chemical Engineering, Edition WILEY-VCH, Weinheim, 2004.

[4] P. Ozil, Plans d'expériences : Méthode de TAGUCHI, Ecole Nationale Supérieure d'Electrochimie et d'Electrométallurgie, INPG, 1997.

[5] J. Goupy, Modélisation par les plans d'expériences, Techniques de l'Ingénieur, R275, 2000.

[6] D. Benoit, Y. Tourbier, S. Germain, Plans d'expériences, construction et analyse, Technique et Documentation, Edition LAVOISIER, Paris 1994.

[7] J. Goupy, Introduction aux plans d'expériences avec application, Edition DUNOD, Paris, 2013.

[8] J. Goupy, Plans d'expériences pour surfaces de réponse, Edition DUNOD, Paris, 1999.

[9] J. Goupy, Pratiquer les plans d'expériences, Edition DUNOD, Paris, 2005.

[10] J. Goupy, Plans d'expériences, Techniques de l'Ingénieur, PE 230, 1997.

[11] K. Hinkelmann, O. Kempthorne, Design and Analysis of Experiments: Advanced Experimental Design, Edition WILEY INTERSCIENCE, New Jersey, 2005.

[12] V. Kafarov. Méthodes cybernétiques et technologie chimique, Edition MIR, Moscou, 1974.

[13] I. Yahiaoui, F. Aissani-Benissad, H. Ait amar, Optimization of silver cementation yield in fixed bed reactor using factorial design and central composite design, The Canadian Journal of Chemical Engineering 6, 1099-1106, 2010.

[14] F. Boudrahem, S. Ziani, F. Aissani-Benissada, Application of Response Surface Methodology Based on Central Composite Design for Optimization of Yellow Bezacryl Sorption on Shoe Soles Waste, Environmental Progress and Sustainable Energy, DOI 10.1002/ep.

[15] J. Goupy, Les plans d'expériences : Les mélanges, Edition DUNOD, Paris, 2000.