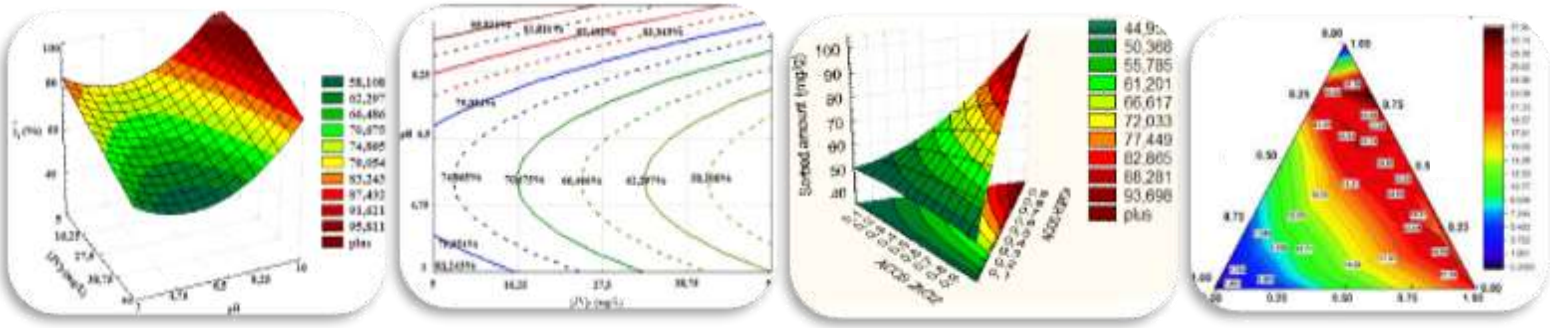
الجمهورية الجزائرية الديمقراطية الشعبية
People's Democratic Republic of Algeria
وزارة التعليم العالي والبحث العلمي
Ministry of Higher Education and Scientific Research
Abderrahmane MIRA University – Bejaia
Faculty of Technology
Department of Process Engineering

Course Notes
For Master's Level Students in Process Engineering

Experimental Designs

Written by:

Dr. MADI-AZEGAGH Katia



University Year 2025/2026

Summary

TABLE OF CONTENTS

Chapitre IV: Response surface plots

Chapitre V: Mixture designs

Introduction

Introduction

Since Antiquity, experimentation has been a fundamental pillar of the scientific approach. However, it was not until the twentieth century that the rigorous planning of experiments known as Design of Experiments, or experimental design emerged as a structured methodology aimed at optimizing experimental work. This methodological approach seeks to extract the maximum amount of relevant information while minimizing the number of experiments required, through careful organization and rational data analysis.

Today, experimental design has become an essential tool across numerous scientific and industrial fields. It not only enables the structured development of experiments, but also supports hypothesis testing, the identification and modeling of interactions between multiple variables, and the determination of optimal operating conditions for a given process.

Unlike fundamental research, which focuses on uncovering the underlying mechanisms of natural phenomena, the methodology of experimental design adopts a pragmatic approach oriented toward optimization, performance improvement, and predictive modeling.
In a context where resources are limited and competitiveness is increasing, the ability to obtain reliable and meaningful results with a reduced number of experiments has become a major challenge. Mastery of experimental design therefore represents a strategic skill for engineers, researchers and technicians.

This document is primarily intended for Master's students in Process Engineering, with particular emphasis on second-year Master's students. It is specifically designed for students specializing in Food Engineering, Chemical Engineering, Pharmaceutical Engineering, and Environmental Process Engineering. This content has been structured to meet the academic and professional needs of these specializations, providing both theoretical foundations and practical applications relevant to their fields of study.

Structured into five chapters, this work provides a progressive introduction to the fundamental principles of experimental design, as well as to the main types of designs commonly used in practice. Its objective is to equip readers with the methodological framework necessary to develop efficient experimental strategies aligned with current research trends and industrial needs.

Chapter I
Introduction and Factorial Designs

I.1 Introduction

Design of Experiments (DOE), also known as experimental design, is statistical tool used to optimize experimental trials. They allow for the planning, analysis, and interpretation of experiments while minimizing the number of trials needed to obtain reliable conclusions. The aim is to understand how different variables influence a given outcome or response.

I.2 What is Experimental Design?

A design of experiments is a structured method used to study the effect of several factors simultaneously. It helps determine how the factors influence a system's response by examining the interactions between these factors. The design of experiment thus allows process optimization while reducing costs and required resources.

Advantages of Experimental Designs (DOE)

- Reduction in the number of trials;
- Ability to study a large number of factors;
- Detection of interactions between factors;
- Identification of optimal conditions;
- Greater accuracy of results;
- Modeling of results.

The use of experimental designs in simple or complex systems allows for the study of various types of functions:

$$y = f(x_i) \qquad \text{(Eq. I.1)}$$

With:

- y: Response of the system
- x_i: Factors or input variables of the system. These factors can be continuous or discrete, qualitative or quantitative.

I.3. Study Domain and Response Surface

The study domain refers to the range of values that the factors of interest can take. The response surface is a graphical or mathematical representation that shows how the response of a system varies as a function of the factors. The goal is often to model this surface in order to identify the optimal conditions for a process or experiment.

I.3.1 Concept of the Study Domain

In an experimental study, the experimenter measures a variable called the response, which is influenced by several factors. These factors are represented on axes and vary between a low level (-1) and a high level (+1). The objective is to understand how changes in these factors affect the measured response. The variation range of each factor includes all the values between these two levels (Fig. I.1).

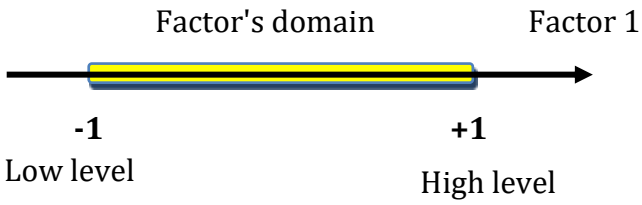


Figure I.1 : Range of variation of a factor

When a second factor is added to the study, it is represented by an axis perpendicular to the first, creating a two-dimensional experimental space. Each factor has a high level and a low level, and their values can be considered as coordinates in this space. An experiment therefore corresponds to a point in this coordinate system, and an experimental design is formed by a set of these points (Fig. I.2).

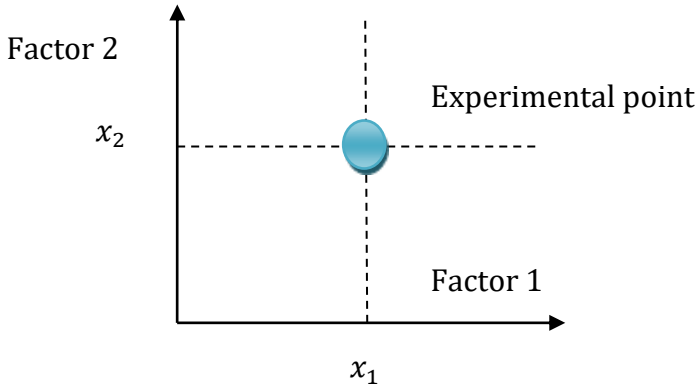


Figure I.2: Experimental zone

The study domain encompasses the ranges of the factors and represents the area of the experimental space selected by the experimenter to conduct the trials. The different experiments carried out are then represented by points distributed within this domain (Fig. I.3).

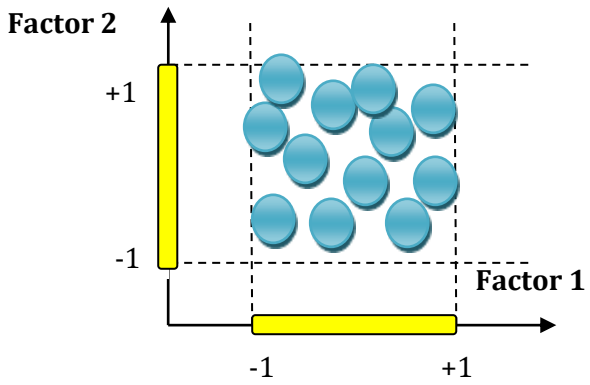


Figure I.3: Experimental points in the study area

I.3.2 Concept of the response surface

In an experimental design, the factor levels define experimental points, while the response is measured at these points. Representing this response requires an additional axis, as a two-factor design operates in a three-dimensional space (Fig. I.4). Each point in the study domain is associated with a response, forming a response surface. The objective is to obtain an accurate representation of this surface while minimizing the number of experiments.

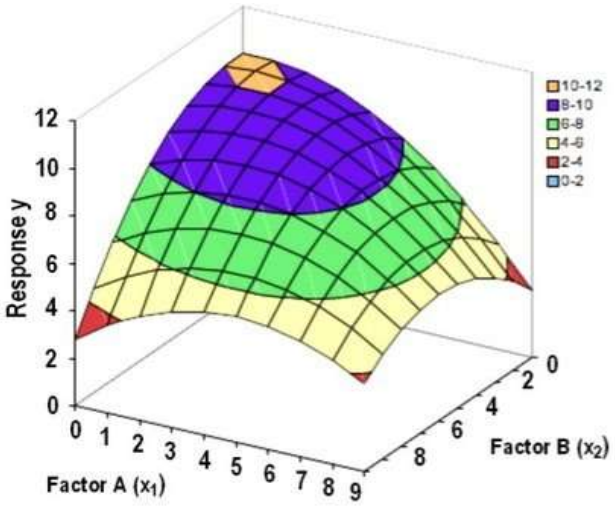


Figure I.4: Response surface

I.4 Effect of a factor

The effect of a factor « x » on the response « y » is assessed by comparing the values of « y » when « x » changes level (from -1 to +1). The corresponding responses, denoted y_1

and y_2 for x = -1 and x = +1 respectively, allow for quantification of this effect. We distinguish between:

- The overall effect as $(y_2 - y_1)$
- The average effect as $(y_2 – y_1)/2$

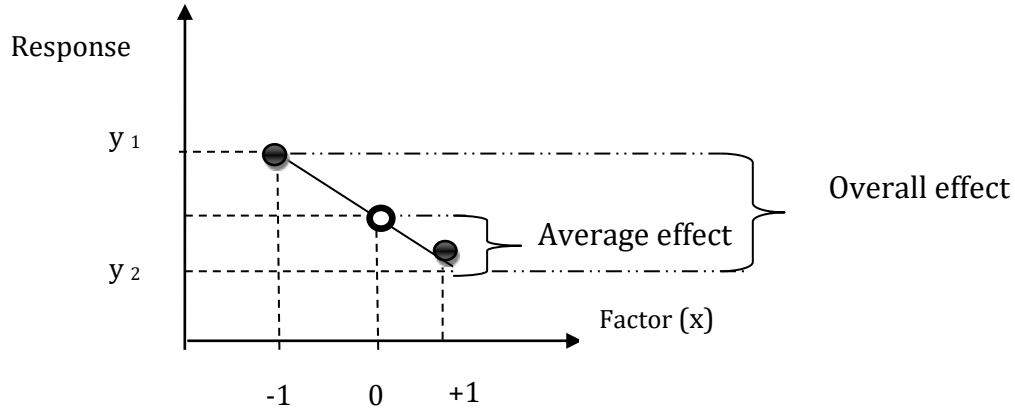


Figure I.5: Illustration of overall effect and average effect

I.5 Notion of interaction

The interaction between factors A and B occurs when the effect of one factor depends on the level of the other. This interaction can be significantly influence the interpretation of data and decision-making in experimental studies. The less parallel the lines representing the effects are, the more likely it is that a significant interaction exists (Fig. I.6).

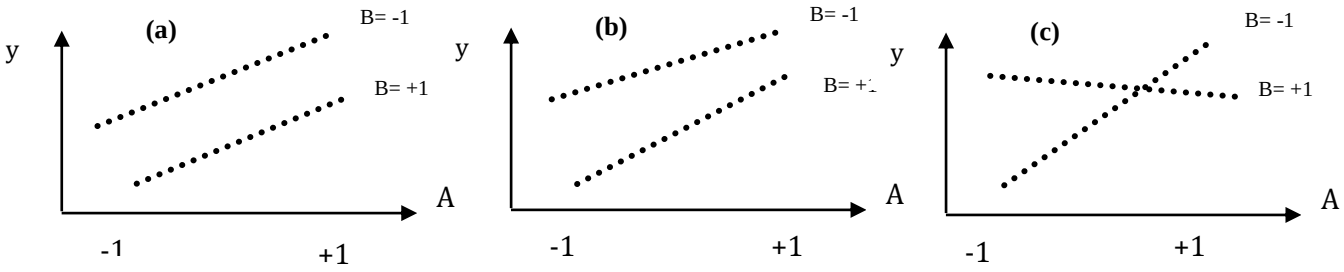


Figure I.6 : Interaction plots:, (a) No interaction between A and B, (b) Weak interaction between A and B and (c) Strong interaction between A and B

I.6 Notion of model and multiple linear regression

Modeling is at the core of experimental design. Its goal is to establish the equation or function that describes how the studied phenomenon varies based on the identified

factors. Thus, each experimental response can expressed by a specific formula, facilitating the analysis and understanding of the results.

The regression equation obtained from the experiment is written in the following form:

$$y = b_0 + \sum_{j=1}^{k} b_j x_j + \sum_{u=1, u \neq j}^{k} b_{uj} x_u x_j \quad \text{(Eq. I.2)}$$

Where

- x_i are the coordinates of the factors z_i expressed as centered and reduced variables (coded variables), (j=1,… ,k) ;
- b_0: Constant term of the regression equation**;**
- b_j : Linear effects ;
- b_{uj} : Interaction effects.

I.7 Full factorial designs 2^k

Full factorial designs 2^k are an effective method for studying the effect of multiple factors on a response while minimizing the number of required experiments. The notation 2^k indicates that each factor has two levels (maximum and minimum), and *k* represents the number of factors studied. For example, with 8 factors, 256 experiments would be required. The objective is to optimally distribute the trials within the experimental domain.

To analyze the data obtained, it is necessary to use dimensionless variables known as centered and reduced variables, or coded variables, instead of real variables. The transformation from real variables to coded variables is done using coding formulas.

I.7.1 Coding formulas

The following relationships are essential for the transition from real variables to coded variables:

$$X_j = \frac{Z_j - Z_j^0}{\Delta Z_j}, \qquad j=1, 2,..., k \quad \text{(Eq. I.3)}$$

$$Z_j^0 = \frac{Z_{jmax} + Z_{jmin}}{2} \quad \text{(Eq. I.4)}$$

$$\Delta Z_j = \frac{Z_{jmax} - Z_{jmin}}{2} \quad \text{(Eq. I.5)}$$

Where :

- $X_1, X_2, \dots, X_k$: Reduced centered variables or coded variables ;
- Z_1, Z_2 ,……, Z_k: Controlled factors (real variables) ;
- $Z_1^0, Z_2^0, \dots, Z_k^0$: Real variables corresponding to the center of the design or sometimes referred to as the fundamental level;
- ΔZ_j : Unit or range of variation along the axis of the Z_j ;
- $Z_{j\,min}$: Minimum value of the real variable;
- $Z_{j\,max}$: Maximum value of the real variable.

Exemple

An experimenter chooses 80 km/h as the low level and 120 km/h as the high level for the factor "car speed."

1. What is the coded value of a speed of 90 km/h?
2. What is the real (uncoded) value of a speed of +0.5 in coded units?

Answer

1) $Z_j^0 = \frac{Z_{jmax} + Z_{jmin}}{2} = \frac{120+80}{2} = 100$ et $\Delta Z_j = \frac{Z_{jmax} - Z_{jmin}}{2} = \frac{120-80}{2} = 20$

$x_j = \frac{Z_j - Z_j^0}{\Delta Z_j} \Rightarrow Z_j = \Delta Z_j\, x_j + Z_j^0 \quad \Rightarrow \quad Z_j = 20\, x_j + 100$

Pour $Z_j = 90\ km/h \quad \Rightarrow \quad x_j = -0.5$

2) $x_j = +0.5 \quad \Rightarrow \quad Z_j = 110$

I.7.2 Calculation of model coefficients

The calculation of the model coefficients is straightforward and is derived from the algebraic properties of the [X] matrix of effects in factorial designs.

$B = [X^T . X]^{-1} . X^T . Y$ (Eq.I.6)

Avec :

- [X] : Matrix of Independent variables;
- $[X^T]$: Transpose matrix of [X];
- $[X^T . X]^{-1}$: Inverse matrix of matrix $[X^T . X]$;
- Y : Observation vector.

I.7.3 Matrix form of multiple linear regression

In matrix form, the initial statistical calculation is presented as follows:

- Column matrix of B coefficients:

$$B = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_k \end{bmatrix}$$

- Observation vector y

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \cdot \\ \cdot \\ \cdot \\ y_N \end{bmatrix}$$

- Matrix of independent variables

$$[X] = \begin{bmatrix} x_{01} & x_{11} & x_{21} & \cdots & x_{k1} \\ x_{02} & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{0N} & x_{1N} & x_{2N} & \cdots & x_{kN} \end{bmatrix}$$

Where x_0 is a fictive variable equal to 1

- Transpose matrix of [X]

$$[x^T] = \begin{bmatrix} x_{01} & x_{02} & x_{03} & \cdots & x_{0N} \\ x_{11} & x_{12} & x_{22} & \cdots & x_{1N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{k1} & x_{k2} & x_{k3} & \cdots & x_{kN} \end{bmatrix}$$

- Matrix product $[X^T X]$:

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

I.8 Application example: Full factorial design 2^4

The objective of this study is to model the degradation process of the pollutant (JV). The operating parameters studied, their respective ranges of variation and the target response (the degradation efficiency of JV, denoted as y) are presented in Tables I.1–3.

Table I.1: Operating parameter values at different levels

Real variables	Reduced center variables	Low level	Central point	High level
		-1	0	+1
Z_1 : Q_v (mL/min)	**x_1**	148.20	224.40	300.60
Z_2 : [JV]$_0$ (mg/L)	**x_2**	16.25	27.50	38.75
Z_3 : pH	**x_3**	4.75	6.50	8.25
Z_4 :[H_2O_2] (mg/L)	**x_4**	100.00	200.00	300.00

Table I.2 : JV degradation efficiency (y (%))

N°	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
y(%)	77.76	78.75	91.39	96.56	50.16	69.70	54.89	79.60
N°	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
y(%)	70.60	73.16	78.20	89.28	37.83	53.51	59.68	67.12

Table I.3: JV degradation yield (y (%)) for tests in the center of the domain

N°	**1**	**2**	**3**
y(%)	69.62	73.56	70.68

I.8.1. First-degree model development

The postulated mathematical model can be written as follows:

$$\hat{y}=b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{14}x_1x_4 + b_{23}x_2x_3 + b_{24}x_2x_4 + b_{34}x_3x_4 + b_{123}x_1x_2x_3 + b_{124}x_1x_2x_4 + b_{234}x_2x_3x_4 + b_{1234}x_1x_2x_3x_4$$

- **Model coefficients**

The model coefficients are given by the following matrix product:

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

The factorial design matrix 2^4 and the model coefficients are presented in Tables I.4 and I.5, respectively:

Table I.4 : Factorial design matrix 2^4

N °	x_0	x_1	x_2	x_3	x_4	x_1x_2	x_1x_3	x_1x_4	x_2x_3	x_2x_4	x_3x_4	$x_1x_2x_3$	$x_1x_2x_4$	$x_1x_3x_4$	$x_2x_3x_4$	$x_1x_2x_3x_4$
1	1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	1
2	1	-1	-1	-1	1	1	1	-1	1	-1	-1	-1	1	1	1	-1
3	1	-1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	1	-1
4	1	-1	-1	1	1	1	-1	-1	-1	-1	1	1	1	-1	-1	1
5	1	-1	1	-1	-1	-1	1	1	-1	-1	1	1	1	-1	1	-1
6	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1
7	1	-1	1	1	-1	-1	-1	1	1	-1	-1	-1	1	1	-1	1
8	1	-1	1	1	1	-1	-1	-1	1	1	1	-1	-1	-1	1	-1
9	1	1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1
10	1	1	-1	-1	1	-1	-1	1	1	-1	-1	1	-1	-1	1	1
11	1	1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	1	1
12	1	1	-1	1	1	-1	1	1	-1	-1	1	-1	-1	1	-1	-1
13	1	1	1	-1	-1	1	-1	-1	-1	-1	1	-1	-1	1	1	1
14	1	1	1	-1	1	1	-1	1	-1	1	-1	-1	1	-1	-1	-1
15	1	1	1	1	-1	1	1	-1	1	-1	-1	1	-1	-1	-1	-1
16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table I. 5: Values of the regression equation coefficients (k=4).

Term Constant	**Linear effects**			
b_0	**b_1**	**b_2**	**b_3**	**b_4**
70.512	-4.339	-11.451	6.578	5.448

Two-factor interaction effects					
b_{12}	**b_{13}**	**b_{14}**	**b_{23}**	**b_{24}**	**b_{34}**
-0.187	0.819	-0.853	-0.317	2.973	0.602

Three-factor interaction effects			
b_{123}	**b_{124}**	**b_{134}**	**b_{234}**
1.784	-1.788	-0.567	-0.986

Four-factor interaction effects
b_{1234}
-1.109

Chapter II
Significance tests and model validation

II.1 Introduction

The objective of this analysis is to determine whether the effects considered significant are genuinely related to the factor or interaction in question, or if they simply result from the natural variability of the phenomenon under study.

II.2 Experimental Errors

Experimental errors are the deviations between observed values and those predicted by the model. They are unavoidable and may be either systematic (bias) or random. Managing these errors is crucial to ensure the reliability of the results. The arithmetic mean is used as a central value, while the standard deviation serves as a measure of data dispersion.

Exemple: Suppose there are four center points within the domain of a full factorial design (Table II.1). Instead of listing these four values individually, a common approach is to summarize them using a central value (typically the mean) along with a measure of variability (such as the standard deviation) to capture the spread around that central value.

Table II.1 : Center Values

N° essais	Center Point (y_i)
1	21.10
2	22.16
3	20.40
4	21.90

- ***Trial Number***

By definition, the arithmetic mean of a set of values is the sum of all the values divided by the number of values. For the values y_i given in Table II.1, the arithmetic mean is equal to:

$$\bar{y} = (21.10 + 22.16 + 20.40 + 21.90)\ /4 = 21.39$$

- ***Standard deviation***

The definition of the standard deviation is a little less straightforward than that of the mean. We will describe its calculation step by step:

- The first step involves determining the deviations from the mean, which correspond to the differences between each individual value and the arithmetic mean of the set.

$$21.10 - 21.39 = -0.29$$

22.16- 21.39 = +0.77

20.40 - 21.39 = -0.99

21.90 - 21.39 = +0.51

Note that the sum of these deviations from the mean is zero. In fact, we've shown that this is always the case:

$$\sum_{i=1}^{i=n}(y_i - \overline{y}) = \sum_{i=1}^{i=n} y_i - \sum_{i=1}^{i=n} \overline{y} = \frac{n}{n}\sum_{i=1}^{i=n} y_i - n\overline{y} = n\,\overline{y} - n\overline{y} = 0 \qquad \text{(Eq. II.1)}$$

Thus, the sum of the deviations from the mean is not a suitable measure of dispersion, as positive and negative deviations cancel each other out. To overcome this, the effect of the signs is neutralized by taking either the absolute value or the square of the deviations.

- These deviations from the mean are therefore squared and added together, resulting in the sum of the squared deviations from the mean.

$$(-0.29)^2 + (0.77)^2 + (-0.99)^2 + (0.51)^2 = 1.917$$

- This sum is divided by the number of values (tests) minus 1 (4 - 1= 3)

$$\frac{1,917}{4-1} = 0.639$$

This quantity is called the variance. It is a fundamental quantity in statistical science. It is found everywhere, and will be put to great use.

$$\text{Ecart-type} = \sqrt{0.639} = 0.799$$

II.3 Significance tests for effects

The significance of the regression equation coefficients is assessed using the Student's t-test (Table A) or the p-value to evaluate the relevance of an effect. Two cases can be distinguished:

* ***Case where each test is repeated m times***

In this case, the mean of the results of the parallel tests is first calculated.

$$\overline{y}_i = \frac{\sum_{u=1}^{m} y_{iu}}{m} \qquad i=1, 2,.,, N \qquad \text{(Eq. II.2)}$$

Where :

* m : Number of repetitions for each test;
* N : Number of experiments.

Then, the sampling variances are determined.

$$S_i^2 = \frac{\sum_{u=1}^{m}(y_{iu}-\bar{y}_i)^2}{m-1} \quad \text{i=1, 2,..., N} \qquad \text{(Eq. II.3)}$$

Then we calculate the reproducibility variance

$$S_{rep}^2 = \frac{\sum_{i=1}^{N} {S_i}^2}{N} \qquad \text{(Eq. II.4)}$$

* ***Case where the center point test is repeated n_0 times***

In this case, the measurement (or reproducibility) variance is estimated using the variance calculated at the center of the experimental domain:

$$S_{rep}^2 = \frac{\sum_{i=1}^{n_0}(y_i-\bar{y}_0)^2}{n_0-1} \qquad \text{(Eq. II.5)}$$

With :

* n_0 : Number of repetitions in the center;
* y_i: Experimental results at the center;
* $\bar{y}_0$: Average over center measurements:

$$\bar{y}_0 = \frac{\sum_{i=1}^{n_0} y_i}{n_0} \qquad \text{(Eq. II.6)}$$

* $(n_0 - 1)$ = f degrees of freedom.

In both cases, the reproducibility variance is essential for estimating the significance of the regression coefficients using the Student's t-test.

$$t_j = \frac{|b_j|}{S_{bj}} \qquad \text{(Eq. II.7)}$$

Where:

* t_j : Follows a normal distribution;
* b_j : Is the jth coefficient of the regression equation
* S_{bj} : Is the quadratic deviation which is defined in the case of a first-degree model by:

$$S_{bj} = \frac{S_{rep}}{\sqrt{N}} \qquad \text{(Eq. II.8)}$$

To verify the significance of each coefficient, the Student's table is used to determine the value of $t_\alpha(f)$ for the chosen significance level α and the number of degrees

of freedom f. Using the part of the table corresponding to a two-tailed test, the test rule is as follows:

* If $t_j > t_\alpha(f)$: Coefficient α_i is significantly different from zero ;
* If $t_j < t_\alpha(f)$: Coefficient α_i is not significant

II.4 Linear model validation

Before accepting the proposed model, it is essential to verify the absence of bias. A model is considered unbiased if it accurately describes the variations in the response according to the selected factors.

* **Search for model bias**

The absence of bias can be verified using the Fisher-Snedecor test (Table B), which involves comparing the residual variance S_{res}^2 with the reproducibility variance S_{rep}^2. The residual variance is estimated by:

$$S_{res}^2 = \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N-l} \quad \text{(Eq. II.9)}$$

Where :

* (N - l) : Degrees of freedom;
* l : Number of significant coefficients in the regression equation;
* y_i : Experimental response;
* $\hat{y}_i$: Response calculated from the model.

The model is considered “unbiased” if the following inequality is verified:

$$F = \frac{S_{res}^2}{S_{rep}^2} < F_{0.95}\ (N-l,\ n_0 - 1) \quad \text{(Eq. II.10)}$$

* **Validation of the regression equation**

Once the absence of bias has been confirmed, the validity of the regression equation can be tested using the Fisher statistic.

$$F = \frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2 / (l-1)}{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 / (N-l)} \quad \text{(Eq. II.11)}$$

Avec : $$\bar{y} = \frac{\sum_{i=1}^{N} y_i}{N} \quad \text{(Eq. II.12)}$$

* $\bar{y}$: Mean of all measurements;
* $\sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2$: Sum of squares due to regression;

* $\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$: Sum of squares of the residuals.

If the calculated F-value is greater than the tabulated value of the Fisher test F_α (f_1, f_2) for the chosen significance level α and degrees of freedom $f_1 = (l - 1)$ et $f_2 = (N - l)$, then the variables selected for modeling have a significant overall effect on y and the regression equation is considered adequate.

II.5 Confidence interval of the model effects

The confidence interval (CI) quantifies the precision of an estimate. In the context of model effects, it provides a range of values within which the estimated effect is likely to lie with a given probability (commonly 95%).

II.6 Analysis of variance and validation of the linear model

II.6.1 ANOVA table

The table below summarizes the analysis of variance (ANOVA), which decomposes the total variability of the data into several sources of variation in order to test the effects of the factors. It presents the sums of squares, the degrees of freedom (df), and the F-statistics.

Table II.2: Analysis of variance

Source of variation	Sum of squares	df	Mean square	F
Regression	SSER	p-1	MSR=$\frac{SSER}{N-1}$	$F_{cal} = \frac{MSR}{MSE}$
Residuals (Error)	SSE	n-p	MSE=$\frac{SSE}{N-p}$	
Total	SST	N-1	MST=$\frac{SST}{N-1}$	

- SSER: Sum of squares due to regression, with (l-1) degrees of freedom, where l is the number of estimated coefficients of the model.
- $SSER = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ (Eq. II.13)
- SSE: Sum of squares of residuals, with (N-l) degrees of freedom, where N is the number of experiments performed.

  $SSE = \sum_{i=1}^{N} \hat{e_i}^2 = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2$ (Eq. II.14)
- SST: Total sum of squares, with (N-1) degrees of freedom.

  $SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$ (Eq. II.15)
- MSR, MSE, MST: Ratio of a sum of squares to its corresponding degree of freedom.

The F-test is then used to compare the calculated F value (F_{cal}) obtained in the previous table with the tabulated F value (F_{tab} (df_1; df_2; α) taken from the Fisher–Snedecor distribution table (Table B), with degrees of freedom $df_1 = l - 1$, $df_2 = N - l$ and a predefined significance level α. If $F_{cal} > F_{tab}$ the linear regression model is considered valid. Otherwise, the model is considered invalid.

II.6.2 Coefficient of determination and correlation

The coefficient of determination (R^2), also referred to as the correlation coefficient, measures the proportion of the total variance explained by the model. A high R^2 indicates that the model adequately explains the variability of the data.

$$R^2 = \frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \quad \text{(Eq. II.16)}$$

$\bar{R}^2$ represents the adjusted multiple regression coefficient. The closer this coefficient of determination is to 1, the better the model fits the data.

$$\bar{R}^2 = R^2 - (1 - R^2)\frac{l-1}{N-l} \quad \text{(Eq. II.17)}$$

II.7 Example of Application

After calculating the coefficients of the regression equation from the example given in Chapter I (Table II.3):

1. Test the significance of the regression coefficients using the Student's t-test.
2. Validate the linear model by checking for model bias and by verifying the regression equation.
3. Plot the residuals diagram.
4. Plot the interactions.

Table II. 3: Values of the regression equation coefficients (k=4).

Constant term	**Linear effects**			
b_0	b_1	b_2	b_3	b_4
70.512	-4.339	-11.451	6.578	5.448
Two-factor interaction effects				
b_{12}	b_{13}	b_{14}	b_{23}	b_{24}	b_{34}
-0.187	0.819	-0.853	-0.317	2.973	0.602
Three-factor interaction effects			
b_{123}	b_{124}	b_{134}	b_{234}
1.784	-1.788	-0.567	-0.986
Four-factor interaction effect			
b_{1234}			
-1.109			

Response

- **Statistical analysis of the regression equation**

a) Verification of the significance of the coefficients

The significance of the coefficients is verified using the Student's test. The tj values are calculated from:

$$t_j = \frac{|b_j|}{\sqrt{S_{bj}{}^2}}$$

With:

S_{bj} : Mean square deviation, defined as:

$$S_{bj} = \sqrt{\frac{S_{rep}^2}{N}}$$

S_{rep}^2 : Reproducibility variance.

The reproducibility variance is estimated from the variance calculated at the center of the experimental domain:

$$S_{rep}^2 = \frac{\sum_{i=1}^{n_0}(y_i - \bar{y}_0)^2}{n_0 - 1} \qquad i = 1, 2, \ldots, n_0$$

We obtain:

$$S_{rep}^2 = 4.157$$

$$S_{bj} = 0.510$$

Table II.4: t_j values corresponding to the different coefficients of the model (k=4).

Constant term	**Linear effects**			
t_0	t_1	t_2	t_3	t_4
138.336	8.513	22.465	12.906	10.689
Two-way interaction effects				
t_{12}	t_{13}	t_{14}	t_{23}	t_{24}	t_{34}
0.367	1.608	1.674	0.622	5.833	1.181
Three-way interaction effects			
t_{123}	t_{124}	t_{134}	t_{234}
3.501	3.508	1.112	1.934
Four-way interaction effect			
t_{1234}			
2.176			

According to the Student's t-table for a two-tailed test (Table A) for a significance level of α = 0.05 and a degrees of freedom f = (n_0 -1) = 2, the tabulated value of the Student's t-test, is 4.30.

Since the values of t_{12}, t_{13}, t_{14}, t_{23}, t_{34}, t_{123}, t_{124}, t_{134}, t_{234} and t_{1234} are lower than the tabulated Student's t value, the corresponding coefficients are not significant and will therefore be removed from the regression equation.

b) Validation Test of the Regression Equation

❖ **Bias Check**

The bias check is performed using the Fisher–Snedecor test by comparing the residual variance to the reproducibility variance.

$$F = \frac{S^2_{rés}}{S^2_{rep}}$$

The value of the residual variance calculated for N = 16 and l = 6 (l is the number of significant coefficients) is:

$$S^2_{rés} = 17.283$$

Which leads to a value of: $$F = \frac{17,283}{4,157} = 4.16$$

The tabulated value of the Fisher–Senedecor test for a significance level of α = 0.05 and degrees of freedom (N - l) = 10 and (n_0 - 1) = 2 is 19.4. Since this value is higher than the calculated value (4.16), the model is therefore considered unbiased.

❖ **Regression significance test**

Once the absence of bias has been confirmed, the validity of the regression equation can be tested using the Fisher statistic :

$$F = \frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2 \Big/ (l-1)}{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \Big/ (N-l)} = 61.512$$

According to the Fisher-Senedecor table (Appendix B), for a significance level of α = 0.05 and degrees of freedom (l - 1) = 5 and (N - l) = 10, the calculated and tabulated Fisher test values are 61.512 and 33.3, respectively. Since the calculated F value is higher than the tabulated value, the regression equation is adequate and the model is valid at 95%. The selected equation for the model is therefore written as:

$$\hat{y} = 70.512 - 4.339\,x_1 - 11.451\,x_2 + 6.578\,x_3 + 5.448\,x_4 + 2.973\,x_2x_4$$

a) Model validation using the center point

The pollutant removal efficiency calculated from the model, $\hat{y}$ (0, 0, 0, 0) = 70.512%, is compared with the value obtained experimentally, y = 71.287%. The absolute difference between these two values is about 0.775%, which is lower than 5%. This result indicates that the obtained model is adequate and accurately simulates the experimental results.

b) Calculation of the coefficient of determination

The calculation of the coefficient of determination is given by:

$$R^2 = \frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} = 0.955$$

The adjusted coefficient is given by:

$$\bar{R}^2 = R^2 - (1 - R^2)\frac{l-1}{N-l} = 0.933$$

These values allow us to state that the chosen model is suitable and appropriate.

c) Residual analysis

By calculating the mean of the residuals:

$$Mean_{residual} = \frac{\sum e_i}{N} = 3.109 \cdot 10^{-13}\ \%$$

According to the residuals plot (shown in Figure II.1), where the points are randomly distributed, and considering the mean value of the residuals ($3{,}109 \cdot 10^{-13}$), it

can be concluded that the obtained linear mathematical model provides a satisfactory representation of the studied phenomenon.

Table II.5: Comparison between predicted and experimental responses for the degradation rate of (JV)

Essais	ŷ(%)	y(%)	Résidus (ei)
1	77.76	77.253	0.508
2	78.75	78.626	0.124
3	91.39	93.978	-2.587
4	96.56	95.351	1.209
5	50.16	48.398	1.763
6	69.70	68.816	0.884
7	54.89	57.985	-3.095
8	79.60	78.404	1.196
9	70.60	68.566	2.034
10	73.16	77.093	-3.933
11	78.20	78.154	0.046
12	89.28	86.566	2.600
13	37.83	39.726	-1.896
14	53.51	52.993	0.517
15	59.68	56.451	3.229
16	67.12	69.718	-2.598

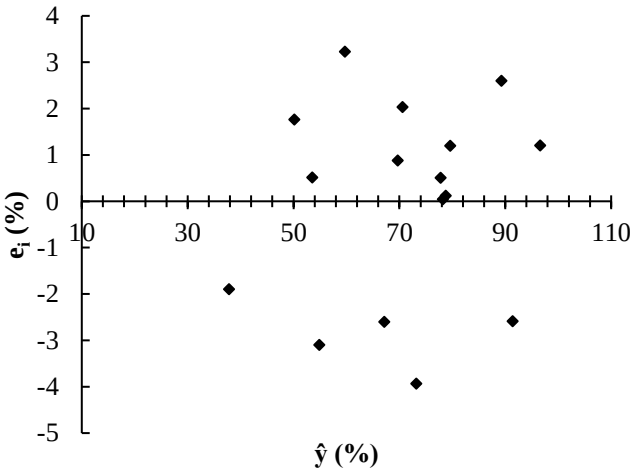


Figure II.1. Residuals plot

- **Use of the developed model**

The analysis of the developed model for decolorization provides information on the most influential parameters. The pH of the solution has a positive effect, with a coefficient of +6.578, as well as the H_2O_2 concentration, with a coefficient of +5.448. In contrast, the flow rate and the initial dye concentration have negative effects, with coefficients of −4.339 and −11.451, respectively.

The interaction analysis based on Figure II.2, allowed us to conclude that the interaction between the initial dye concentration and the H_2O_2 concentration is weak.

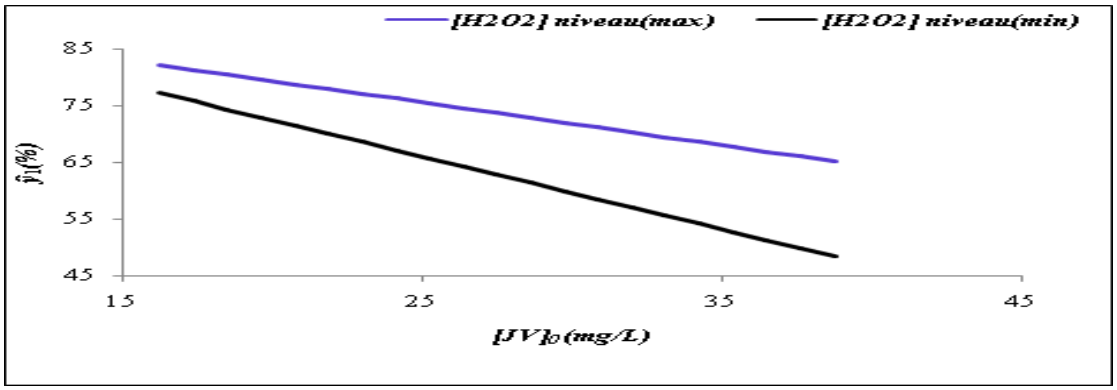


Figure II.2. Interaction between the initial dye concentration and the H_2O_2 concentration.

Table A. Student's t-distribution table

Unilateral Bilateral	**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 B. Fisher-Senedecor table for p = 0.95

f_1: degrees of freedom in the 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 III
Fractional factorial designs

III.1 Introduction

Fractional Factorial Designs (FFD) are tools used in experimental design to reduce the number of experimental runs while retaining a maximum amount of information about the studied factors. Unlike full factorial designs, which require a large number of experiments to investigate all possible combinations of factors, fractional factorial designs make it possible to examine the main effects with a reduced number of runs. This is particularly useful when the number of factors is high and when each experiment is costly or time-consuming.

III.2 Two-level fractional factorial designs 2^{k-q}

At the end of an experiment using a fractional factorial design, a system with more unknowns (p) than equations (n) is obtained, making it impossible to solve directly. To overcome this issue, some coefficients are grouped together so that only n unknowns remain. These groupings are called contrasts or aliases. The coefficients are then said to be aliased.

A fractional design has: $N= 2^{k-q}$ experimental runs.

$q = 1$: $N = 2^{k-1}$ tests $\Rightarrow 2^1$ fewer experimental runs than the corresponding full factorial design;

$q = 2$: $N = 2^{k-2}$ tests $\Rightarrow 2^2$ fewer experimental runs than the corresponding full factorial design;

⋮ ⋮

⋮ ⋮

$q = q$: $N = 2^{k-q}$ tests $\Rightarrow 2^q$ fewer experimental runs than the corresponding full factorial design.

Where :

k : number of factors;

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

q : indicates that the number of experimental runs in the design has been divided by 2^q.

III.3 Designing a fractional plan

The design of a fractional factorial plan is based on a solid understanding of aliasing theory, which is essential for analyzing and interpreting the results. The study of this theory is organized around five main aspects:

III.3.1 Definition of contrasts

When interpreting the results of a fractional design, the same calculation rules apply as for full designs.

III.3.2 Interpretation hypotheses

To analyze the results of a fractional factorial design, Hypothesis 1 must be supplemented with additional assumptions:

- ***Hypothesis 1:*** Third-order interactions and all higher-order interactions can be neglected.
- ***Hypothesis 2:*** All coefficients aliased within a weak (negligible) contrast are themselves weak (negligible).
- ***Hypothesis 3:*** If two contrasts are strong, their interaction may also be strong and should be considered carefully.
- ***Hypothesis 4:*** If two contrasts are weak, their interaction is assumed to be weak as well.
- ***Hypothesis 5:*** A weak main factor combined with a strong main factor most often, but not always, produces a weak interaction.

The hypotheses presented here are very often verified; however, there are cases where they may not hold. It is always possible to adopt alternative assumptions depending on the problem under study and the associated risks. For a proper analysis of the results, it is advisable to clearly specify the interpretation assumptions that have been adopted.

III.3.3 Box calculation

Box calculation allows for quickly determining how effects and interactions are aliased within the contrasts. It is applied to fractional factorial designs with two fixed levels at –1 and +1.

- **Box notation:** In Box notation, the number (in bold) represents the column of signs for factor 1, with the signs arranged according to Yates' order

  For a 2^2 design, we have : Pour un plan 2^2, on a :

$$\mathbf{1} = \begin{bmatrix} -1 \\ +1 \\ -1 \\ +1 \end{bmatrix}$$

The factor 2 column will be designated by:

$$\mathbf{2} = \begin{bmatrix} -1 \\ -1 \\ +1 \\ +1 \end{bmatrix}$$

Columns of minus and plus signs are introduced:

$$\mathbf{I} = \begin{bmatrix} +1 \\ +1 \\ +1 \\ +1 \end{bmatrix} \qquad -\mathbf{I} = \begin{bmatrix} -1 \\ -1 \\ -1 \\ -1 \end{bmatrix}$$

- **Rules to remember**

You just need to remember the following rules for Box calculation. To simplify matters, the multiplied signs have been omitted from the formulas:

- ✓ **Rule 1** : Commutativity : **1 * 2 = 2 * 1**
- ✓ **Rule 2** : Multiplying a column by itself: **1 * 1 = I** et **2 * 2 = I**
- ✓ **Rule 3** : Multiplying a column by I or – I :

  1* I = 1 et **2 * I = 2**

  1 * (–I) = **–1** et **2 * (–I) = –2**

  I * I = I

III.3.4 Equivalence relation

- **Basic design**

Consider a 2^3 factorial design and its postulated model:

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

The term x_1x_2 is the product of the levels of factors 1 and 2. This term represents the 1–2 interaction. A column of signs corresponding to the 1–2 interaction can therefore be constructed by multiplying, according to Box calculation, the columns of factor 1 and factor 2. The same procedure can be applied to the 1–3, 2–3, and 1–2–3 interactions. A column of plus signs can also be added to include the constant in the calculations. This results in a table, called the base matrix or base design, which contains 8 rows and 8 columns (Table III.1)

Table III.1: Basic matrix of a 2^3

Test N°	**I**	**1**	**2**	**3**	**1*2**	**1*3**	**2*3**	**1*2*3**
1	+1	-1	-1	-1	+1	+1	+1	-1
2	+1	+1	-1	-1	-1	-1	+1	+1
3	+1	-1	+1	-1	-1	+1	-1	+1
4	+1	+1	+1	-1	+1	-1	-1	-1
5	+1	-1	-1	+1	+1	-1	-1	+1
6	+1	+1	-1	+1	-1	+1	-1	-1
7	+1	-1	+1	+1	-1	-1	+1	-1
8	+1	+1	+1	+1	+1	+1	+1	+1

This base matrix corresponds to the matrix used to calculate the effects and interactions of a full factorial design. However, for a fractional design, only half of the experiments from this matrix are used.

Let us divide this base matrix into two fractions: a half-design consisting of experiments N° 5, 2, 3, and 8 (darker lines in Table III.2), and another half-design consisting of experiments N° 1, 6, 7, and 4.

Table III.2 : Effect matrix sorted according to the fractional factorial design (FFD)

Test N°	**I**	**1**	**2**	**3**	**1*2**	**1*3**	**2*3**	**1*2*3**	
5	+1	+1	-1	-1	-1	-1	+1	+1	1st half-design
2	+1	-1	+1	-1	-1	+1	-1	+1	
3	+1	-1	-1	+1	+1	-1	-1	+1	
8	+1	+1	+1	+1	+1	+1	+1	+1	
1	+1	-1	-1	-1	+1	+1	+1	-1	2nd half-design
6	+1	+1	+1	-1	+1	-1	-1	-1	
7	+1	+1	-1	+1	-1	+1	-1	-1	
4	+1	-1	+1	+1	-1	-1	+1	-1	

First, let us focus on the upper half-fraction (Table III.2), represented by the darker plan. It contains eight columns of four signs that are identical in pairs. Using Box notation,

we can state, for example, that the column of four signs corresponding to factor 1 is identical to that of the 23 interaction, that is:

$$1 = 2*3$$

We have seen that, in this half-fraction, factor 1 is confounded with the 23 interaction, and that the contrast l_1 calculated using column 1 (or column 23) corresponds to the sum of the coefficients a_1 and a_{23}:

$$l_1 = l_{23} = a_1 + a_{23}$$

It can be observed that the aliased coefficients within a contrast are those that share identical sign columns in the considered half-fraction. This is a general rule: A contrast is the algebraic sum of the coefficients that have the same sign columns. That is, if we have 1=23 in Box notation, the contrast calculated using column 1 is given by $l_{23} = a_1 + a_{23}$, and the contrast calculated using column 23 is $l_{23} = a_1 + a_{23}$. This is called the equivalence relation. It is valid in both directions and forms the basis of alias theory.

An examination of Table III.2 also shows that:

2 = 1*3 is equivalent to $l_2 = l_{13} = a_2 + a_{13}$

3 = 1*2 is equivalent to $l_3 = l_{12} = a_3 + a_{12}$

I = 1*2*3 is equivalent to $l_0 = 123 = a_0 + a_{123}$

These relations are important because they make it possible to identify which effects and interactions are confounded (aliased) within each contrast. To determine the alias structure, one could write the base design, divide it into fractions, and then examine which sign columns are identical. However, although this method is useful for understanding, it is too long and complex. Fortunately, by using Box's notation and alias generators, the alias structure can be easily and efficiently determined.

III.3.5 Alias generators

a) Upper Plane Alias Generator

Consider the upper half-plane of Table III.2. The two columns of plus signs allow us to write in Box notation:

$$\mathbf{I} = 1*2*3$$

By multiplying both sides of this equation by 1 and applying Box's multiplication rules, we obtain the identical columns of the upper half-design:

- ✓ Multiply both sides by 1:

$$\mathbf{1*I} = 1*1*2*3$$

- ✓ Application of Rule 2:

$$\mathbf{1*I} = I*2*3$$

- ✓ Application of Rule 3:

$$\mathbf{1} = 2*3$$

By multiplying both sides of the equation I = 123 successively by 2 and then by 3, we obtain the following two equations:

$$\mathbf{2} = 1*3$$

$$\mathbf{3} = 1*2$$

The relation I=1*2*3 therefore makes it possible to identify all the columns that are identical in the upper half-fraction. By knowing these equalities and applying the equivalence relation, we can determine how the coefficients are aliased within the contrasts.

The relation I=1*2*3 I is called the alias generator.

a) Alias generator for the lower design

By considering the lower half-fraction (Table III.2), it can be verified that the columns correspond pairwise but with opposite signs: to each column with positive signs corresponds a column with negative signs of the 123 interaction. Using Box notation, we can write:

$$\mathbf{I} = -1*2*3$$

This is the alias generator for the lower half-design.

Multiplying both sides of the previous relation by 1 and taking into account Box's multiplication rules, we have:

$$\mathbf{1*I} = 1*(-1*2*3)$$

$$\mathbf{1} = -I*2*3$$

$$\mathbf{1} = -2*3$$

Column 23 of the interaction has many signs opposite to those in column 1.

We find the contrast structure of column 1 by applying the equivalence relation:

$$l'_1 = -\,l'_{23} = a_1 - a_{23}$$

The contrasts calculated using the lower half-fraction correspond to differences between coefficients.

It can also be shown that

2 = –1*3 that is $l'_2 = -l'_{13} = a_2 - a_{13}$

3 = –1*2 that is $l'_3 = -l'_{12} = a_3 - a_{12}$

I = –1*2*3 that is $l'_0 = -l'_{123} = a_0 - a_{123}$

III.4 Application example

Consider a study on the removal of a pharmaceutical pollutant (KTP) using the electrocoagulation process. The parameters to be optimized are: pH (x_1), initial KTP concentration (x_2), and current density (x_3). Table III.3 presents the matrix of the full 2^3 factorial design.

Table III.3 : PFC effects matrix

Test 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	39.29
2	+1	+1	-1	-1	-1	-1	+1	+1	47.52
3	+1	-1	+1	-1	-1	+1	-1	+1	99.64
4	+1	+1	+1	-1	+1	-1	-1	-1	99.50
5	+1	-1	-1	+1	+1	-1	-1	+1	56.98
6	+1	+1	-1	+1	-1	+1	-1	-1	62.55
7	+1	-1	+1	+1	-1	-1	+1	-1	98.26
8	+1	+1	+1	+1	+1	+1	+1	+1	98.82

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

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

Table III.3: Effect matrix sorted according to the FFD

Test 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	47.52	1st half-design
3	+1	-1	+1	-1	-1	+1	-1	+1	99.64	
5	+1	-1	-1	+1	+1	-1	-1	+1	56.98	
8	+1	+1	+1	+1	+1	+1	+1	+1	98.82	
1	+1	-1	-1	-1	+1	+1	+1	-1	39.29	2nd half-design
4	+1	+1	+1	-1	+1	-1	-1	-1	99.50	
6	+1	+1	-1	+1	-1	+1	-1	-1	62.55	
7	+1	-1	+1	+1	-1	-1	+1	-1	98.26	

2. Select either the upper or the lower half-fraction and perform the study based on this part (i.e., carry out the experiments corresponding to the selected fraction).
3. Identify the aliased columns of the new effect matrix that exhibit identical sequences of signs (+1) and (–1) in pairs. In the table above, column x_0 and column x_{123} are identical, which is equivalent to writing 0=1*2*3.

   Similarly, columns x_1 and x_{23}, columns x_3 and x_{12}, and columns x_2 and x_{13} are identical, which leads to:

   $$1=2*3, \quad 2=1*3 \text{ et } \quad 3=1*2$$

4. Determine the contrasts (apparent effects resulting from the fact that effects are confounded pairwise).

   a) Upper half-fraction: The alias generator is column 1*2*3. The defining relation is:

   $$0=I=1*2*3$$

According to Box, the calculation is based on the following relations:

✓ Multiplying a column by itself gives a column of positive signs (+); thus:

$$I : 1*1= 1^2= I$$

✓ Multiplying a column by the column of positive signs III reproduces the original column:

$$1*I= I*1=1$$

✓ Let us multiply both sides of the defining relation I = 1*2*3 by 1. We obtain:

$$1*I = 1*1*2*3 = 1*1*2*3 = 1^2*2*3$$

By simplifying according to the above rules (Box algebra), we obtain:

$$1 = 2*3$$

This means that the main effect 1 is aliased with the interaction 2*3.

✓ Applying the same procedure to all factors, we obtain:

2 = 1*3 (main effect 2 is aliased with interaction 1*3)

3 = 1*2 (main effect 3 is aliased with interaction 1*2)

Once the aliases are identified, the contrasts Li are then determined.

$$l_1 = l_{23} = a_1 + a_{23}$$

$$l_2 = l_{13} = a_2 + a_{13}$$

$$l_3 = l_{12} = a_3 + a_{12}$$

$$l_0 = l_{123} = a_0 + a_{123}$$

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

Table III.4 : FFD effect matrix

Test N°	x_0	x_1	x_2	x_3	y(%)
2	+1	-1	-1	-1	47.52
3	+1	+1	+1	-1	99.64
5	+1	+1	-1	+1	56.98
8	+1	-1	+1	+1	98.82

Using this reduced matrix X_{sup}, the contrasts are calculated using the formula:

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

The resulting values are listed in Table III.5.

Table III.5: Contrast values of the model for the upper half-fraction.

Apparent effects	l_0	l_1	l_2	l_3
values	75.74	2.57	23.49	2.16

The model obtained after calculation is then written as:

$$\hat{y}(\%) = 75.74 + 2.57x_1 + 23.49x_2 + 2.16x_3$$

b) Lower half-fraction: In the lower half-fraction, column 1*2*3 contains only (-1) values, unlike column 0=I, which contains only (+1) values. The defining relation can thus be written as:

$$0=I=-1*2*3.$$

✓ Multiplying both sides of the defining relation I=−123I by 1, we obtain:

$$1\times I = -1\times 1*2*3 = -1\times 1\times 2*3 = -1^2\times 2*3$$

By simplifying according to the Box algebra rules mentioned above, we get:

$$1 = -2*3$$

This means that the main effect 1 is aliased with the interaction -2*3.

Table III.6: FFD effect matrix

Test N°	x_0	x_1	x_2	x_3	y(%)
1	+1	+1	-1	-1	39.29
4	+1	-1	+1	-1	99.50
6	+1	-1	-1	+1	62.55
7	+1	+1	+1	+1	98.26

✓ Applying the same procedure to all factors, we obtain:

2=−1*3 (main effect 2 is aliased with interaction -1*3)

3=−1*2 (main effect 3 is aliased with interaction -1*2)

As with the upper half-fraction, the contrast values depend on the actual effects of the full factorial design. We can write:

$$l_1 = -l_{23} = a_1 - a_{23}$$

$$l_2 = -l_{13} = a_2 - a_{13}$$

$$l_3 = -l_{12} = a_3 - a_{12}$$

$$l_0 = -l_{123} = a_0 - a_{123}$$

Thus, the system obtained after reducing the number of unknowns is as follows:

The contrasts calculated for this half-desig are listed in Table III.7.

Table III.7 : Contrast values of the model for a lower half-plane

Apparent effects	l_0	l_1	l_2	l_3
values	75.90	-6.12	23.98	5.50

The model obtained after calculation is then written as:

$$\hat{y}(\%) = 75.90 - 6.12x_1 + 23.98x_2 + 5.50x_3$$

Note: n the lower half-fraction, column 123 contains only (-1) values, unlike column I, which contains only (+1) values. This gives the defining relation:

$$I = -1*2*3$$

From this, all aliases can be determined by writing, for example:

$$1*I = -1*1*2*3 = -1^2*2*3 \Rightarrow 1 = -2*3.$$

Once the contrast values are determined, the statistical analysis proceeds in the same way as presented in Chapter II, in order to validate the proposed mathematical model before its interpretation.

III.5 Other Designs: Plackett-Burman and Taguchi Designs

The matrices used to construct Plackett-Burman designs are Hadamard matrices. These are matrices with 4, 8, 12, 16, 20, 24, 28, 32, 36, etc., rows. They allow experiments to be conducted with an intermediate number of runs compared to classical factorial designs, which only have 2^k runs (4, 8, 16, 32, etc.).

Plackett-Burman designs are generally saturated: they use the minimum number of runs necessary to estimate all main effects. In this case, the mathematical model is a no-interaction model, similar to that used in Koshal designs.

$$y = a_0 + \sum a_i x_i$$

Chapter IV
Response surface plots

IV.1 Introduction

The previously studied designs, with two levels per factor, are primarily used for screening with first-order models. However, when more precise modeling is required, second-order polynomial models must be used. It is in this context that response surface designs come into play, including central composite designs, Box-Behnken designs and Doehlert designs which allow a more detailed exploration of the effects and interactions between factors.

IV.2 Concept of response surface and iso-response curves

In an experimental design, the experimental points define a response surface that represents the evolution of the response as a function of the factors. This surface, visualized in three dimensions for two factors, can be projected as iso-response curves, which connect points with the same response value. These curves facilitate the analysis of results and the identification of optimal conditions. The objective is to model this surface accurately while minimizing the number of experimental runs (Fig. IV.1).

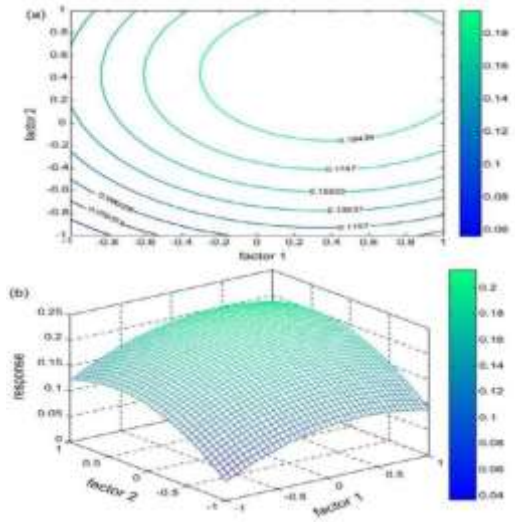


Figure IV.1: Formation of the response surface and iso-response curves from experimental points.

Response surface analysis is based on the use of second-degree polynomial models, generally expressed in the following form:

$$\hat{y}=b_0+\sum_{j=1}^{k} b_j x_j+\sum_{u=1,u\neq j}^{k} b_{uj}\ x_u x_j+\sum_{j=1}^{k} b_{jj} x_j^2 \qquad \text{(Eq.IV.1)}$$

Where

- $\hat{y}$: Response predicted by the model;

- x_j : Coordinates of the Zj factors expressed as coded centered variables (j=1, …, k) ;
- b_0 : Constant term of the regression equation;
- b_j : Linear effects;
- b_{uj} : Interaction effects;
- b_{jj} : Quadratic effects.

IV.3 Designs for studying quadratic models

IV.3.1 Centered composite design

A central composite design is used in the construction of a response surface. It allows for the development of second-order mathematical models and is applied to continuous variables. The central composite design consists of conducting the runs of a factorial design, supplemented by experiments at the center of the study domain and axial (star) runs (Fig. IV.2).

The number of levels in a central composite design is five per factor: the central point (0), the two factorial design levels (-1, +1), and the two axial (star) points (-ω, +ω). The total number of trials (N) to be performed is the sum of : N= n_f + 2k + n_0

- n_f : Complete factorial design runs;
- 2k : Star runs on axes at a distance ω from the center of the domain;
- n_0 : Runs at the center of the domain.

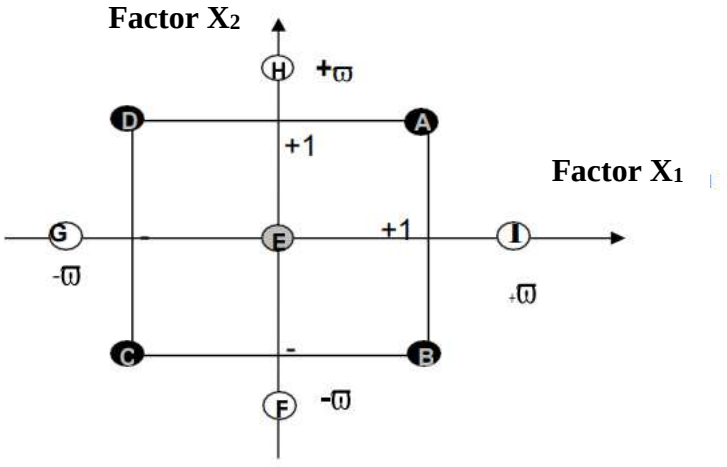


Figure IV.2: Central composite design for two factors (k=2)

IV.3.1.1 Properties of composite planes

a) Postulated mathematical model

The mathematical model constructed using composite designs is a second-degree polynomial with interactions. In general, the interactions are of second order. For two factors (k = 2), the regression equation is written as:

$$\boldsymbol{y} = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 + b_{11} x_1^2 + b_{22} x_2^2 + \varepsilon \qquad \text{(Eq.IV.2)}$$

b) Calculation matrix

For a two-factor composite design, 12 experiments are therefore necessary to determine the values of the 6 coefficients of the model equation. Thus, the test matrix [X] is as follows:

$$\boldsymbol{X} = \begin{bmatrix} +1 & -1 & -1 & +1 & +1 & +1 \\ +1 & +1 & -1 & -1 & +1 & +1 \\ +1 & -1 & +1 & -1 & +1 & +1 \\ +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & 0 & 0 & 0 & 0 & 0 \\ +1 & 0 & 0 & 0 & 0 & 0 \\ +1 & -\omega & 0 & 0 & +\omega^2 & 0 \\ +1 & +\omega & 0 & 0 & +\omega^2 & 0 \\ +1 & 0 & -\omega & 0 & 0 & +\omega^2 \\ +1 & 0 & +\omega & 0 & 0 & +\omega^2 \\ +1 & 0 & 0 & 0 & 0 & 0 \\ +1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

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

$$S_{bj}^2 = C_{jj}.S_{rep}^2 \qquad \text{(Eq.IV.3)}$$

The coefficients are then estimated in the same way as for the full factorial design. The values of the parameter ω and the number n_0 of points at the center depend on the number of factors k in the basic factorial design and the optimality criterion satisfied by the design.

c) Optimality criteria

For a given model type, the objective is to determine the optimal placement of the experimental points for which the error on the predicted responses is as low as possible.

* **Iso-variance criterion by rotation**

For the composite design to satisfy this criterion, the star points must be placed at a distance « ω » equal to the fourth root of the number of points in the factorial design. In the English-language literature, this is referred to as the rotatability criterion. « ω » is calculated using the following relationship:

$$\omega = n_f^{1/4} \qquad \text{(Eq.IV.4)}$$

* **Criterion of near orthogonality**

The objective of this criterion is to get as close as possible to an orthogonal situation. In the case of central composite designs, the submatrix obtained by removing the first row and the corresponding first column (associated with the constant and squared terms) of the matrix $[X^T \cdot X]^{-1}$ is diagonal if the chosen ω satisfies the following condition:

$$\omega = \left(\frac{n_f\left(\sqrt{N}-\sqrt{n_f}\right)^2}{4}\right)^{1/4} \qquad \text{(Eq.IV.5)}$$

The corresponding dispersion matrix $[X^T \cdot X]^{-1}$ is nearly orthogonal, and therefore the criterion of near orthogonality is satisfied.

* **Uniform accuracy criterion**

If the variance function is practically constant within a sphere centered at the same point as the experimental domain, and iso-variance is already ensured, the experimental design exhibits uniform precision properties. Its purpose is to guarantee the same level of precision for the predicted response throughout the entire domain. Table IV.1 summarizes the values of ω and n_0 according to the different optimality criteria.

Table IV.1: Values of ω and n_0 according to the properties sought for the composite design

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
$N_e = 2k$	4	6	8	10	10	12	12
➢ **Iso-variance criterion by rotation**	≥1	≥1	≥1	≥1	≥1	≥1	≥1
➢ **Uniform Precision**	5	6	7	10	6	15	9
➢ **Orthogonality**	8	12	12	17	10	24	15
ω	1.41	1.68	2	2.38	2.00	2.83	2.38

Avec :

- N_f: Number of runs in the factorial design;
- N_e: Number of axial (star) runs;
- n_0 : Number of center-point runs;
- k : Number of factors.

IV.3.2 Box-Behnken design

Box-Behnken designs, proposed in 1960, allow the construction of second-order models with factors at three levels (−1, 0, +1). They are easy to implement and offer flexibility through their sequential nature, allowing factors to be added without losing the previous runs. The experimental points are placed at the center of the edges, and faces of the cube, distributing them at equal distances from the center of the study domain, i.e., on a sphere or hypersphere.

For three factors (K=3), the design includes N_f=2K (K-1) =12 points on the edges of the cube and generally n_0=3 center points, giving a total of N=2K (K-1) + n_0 =15 runs (Fig. IV.3). Adding an extra center point allows the design to achieve near-orthogonality.

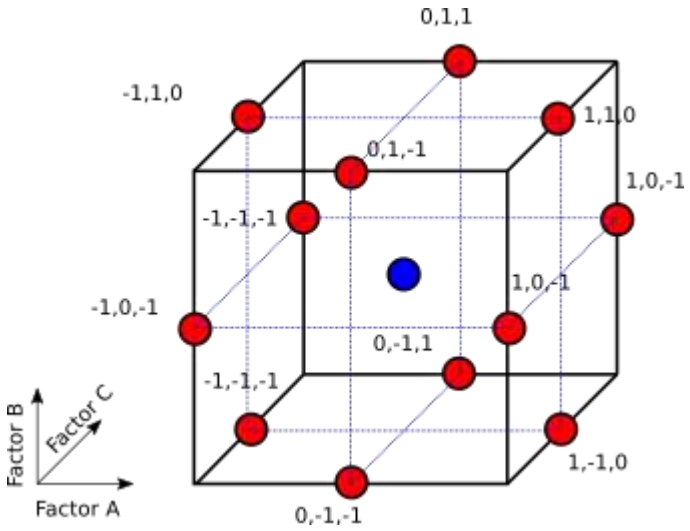


Figure IV.3: Box-Behnken design for three factors

This design includes twelve runs, to which one or more center points can be added. Table IV.2 presents these twelve runs completed with three center points within the

experimental domain. The structure of the Box-Behnken design matrix gives it rotational invariance, reflecting a spherical symmetry of the experimental points.

Table IV.2: Box-Behnken design for three factors

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

The classical mathematical model derived from the general model (Eq. IV.6) and representing the Box-Behnken design with three factors is written as:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2$$

(Eq. IV.6)

IV.4 Application Example: Response surface designs

To conduct a study based on a central composite design with four factors and uniform precision, we need to perform $N=N_f+n_0+N_e=2^4+3+2*4=27$ runs. The mathematical model associated with this design is a model with M = 15 coefficients, which can be written in the following form:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{14}x_1x_4 + b_{23}x_2x_3 + b_{24}x_2x_4 + b_{34}x_3x_4 + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 + b_{44}x_4^2$$

The factor levels and their corresponding coded variables are listed in Table IV.3.

Table IV.3: Operating parameter values at different levels

Real Variables	Coded Variables	Low Level		Center Point	High Level	
		-2	-1	0	+1	+2
Z_1 : Q_v (mL/min)	x_1	72.00	148.20	224.40	300.60	377.40
Z_2 : $[JV]_0$ (mg/L)	x_2	5.00	16.25	27.50	38.75	50.00
Z_3 : pH	x_3	3.00	4.75	6.50	8.25	10.00
Z_4 : $[H_2O_2]$ (mg/L)	x_4	0.00	100.00	200.00	300.00	400.00

The values -2 and +2 correspond, respectively, to the low and high axial (α) for a number of factors k=4.

To develop the second-order model, the results of the sixteen runs from the full factorial design with interactions, as well as the three center-point runs (example given in Chapter I), are retained and supplemented with the eight additional runs of the star (axial) design. The experimental matrix for the second-order design is presented in Table IV.4.

Table IV.4 : Experimental matrix of a central composite design.

Runs N°	Q_v (mL/min)	$[JV]_0$ (mg/L)	pH	$[H_2O_2]$ (mg/L)	x_0	x_1	x_2	x_3	x_4	y(%)
01	148.2	16.25	4.75	100	+1	-1	-1	-1	-1	77 .76
02	148.2	16.25	4.75	300	+1	-1	-1	-1	+1	78.75
03	148.2	16.25	8.25	100	+1	-1	-1	+1	-1	91.39
04	148.2	16.25	8.25	300	+1	-1	-1	+1	+1	96.56
05	148.2	38.75	4.75	100	+1	-1	+1	-1	-1	50.16
06	148.2	38.75	4.75	300	+1	-1	+1	-1	+1	69.70
07	148.2	38.75	8.25	100	+1	-1	+1	+1	-1	54.89
08	148.2	38.75	8.25	300	+1	-1	+1	+1	+1	79.60
09	300.6	16.25	4.75	100	+1	+1	-1	-1	-1	70.60
10	300.6	16.25	4.75	300	+1	+1	-1	-1	+1	73.16
11	300.6	16.25	8.25	100	+1	+1	-1	+1	-1	78.20
12	300.6	16.25	8.25	300	+1	+1	-1	+1	+1	89.28
13	300.6	38.75	4.75	100	+1	+1	+1	-1	-1	37.83
14	300.6	38.75	4.75	300	+1	+1	+1	-1	+1	53.51
15	300.6	38.75	8.25	100	+1	+1	+1	+1	-1	59.68
16	300.6	38.75	8.25	300	+1	+1	+1	+1	+1	67.12
17	224.4	27.50	6.50	200	+1	0	0	0	0	69.62
18	224.4	27.50	6.50	200	+1	0	0	0	0	73.56
19	224.4	27.50	6.50	200	+1	0	0	0	0	70.68
20	72.0	27.50	6.50	200	+1	-2	0	0	0	63.39
21	377.4	27.50	6.50	200	+1	+2	0	0	0	56.86

22	224.4	5.00	6.50	200	+1	0	-2	0	0	91.21
23	224.4	50.00	6.50	200	+1	0	+2	0	0	44.93
24	224.4	27.50	3.00	200	+1	0	0	-2	0	79.13
25	224.4	27.50	10.00	200	+1	0	0	+2	0	92.54
26	224.4	27.50	6.50	0	+1	0	0	0	-2	32.76
27	224.4	27.50	6.50	400	+1	0	0	0	+2	65.91

- ***Results Analysis***

Calculation of the model coefficients concerning the degradation rate of JV. The estimation of the regression equation coefficients is performed using the following matrix product:

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

The results of this calculation are compiled in the following table:

Table IV.5 : Values of the coefficients b_j of the regression equation.

Interaction and quadratic effects									
b_{12}	b_{13}	b_{14}	b_{23}	b_{24}	b_{34}	b_{11}	b_{22}	b_{33}	b_{44}
-0.187	0.819	-0.853	-0.317	2.973	0.602	-2.012	-0.026	4.416	-4.709

Constant term	Linear effects			
b_0	b_1	b_2	b_3	b_4
71.287	-3.437	-11.490	5.503	6.395

- ***Statistical Analysis of the Regression Equation***
- ***Verification of the Significance of the Coefficients***

The reproducibility variance S^2_{rep} is the same as that calculated for the factorial design, $S^2_{rep} = 4.157$. To determine the variance of the model coefficients, it is sufficient to multiply the diagonal elements C_{jj} of the dispersion matrix $[X^T.X]^{-1}$ by this variance: $S^2_{bj} = C_{jj}.S^2_{rep}$

The values of t_j from the Student's t-test are compiled in Table IV.6.

Table IV.6 : t_j Values

Constant term	Linear effects			
t_0	t_1	t_2	t_3	t_4
60.560	8.259	27.609	13.222	15.365

Interaction and quadratic effects									
t_{12}	t_{13}	t_{14}	t_{23}	t_{24}	t_{34}	t_{11}	t_{22}	t_{33}	t_{44}
0.367	1.608	1.674	0.622	5.833	1.181	4.558	0.058	10.003	10.669

For a significance level of α = 0.05 and a degree of freedom f = (n_0-1) = 2, the tabulated value of the Student's t-test $t_\alpha(f)$ is 4.3. Since the values of t_{12}. t_{13}. t_{14}. t_{23}. t_{34} and t_{22} are all less than $t_\alpha(f)$, these coefficients are therefore removed from the regression equation.

- **Bias investigation**

The calculated value of the residual variance for N = 27 and l = 9 is given by:

$$S^2_{res} = 22.488$$

Which leads to:

$$F = \frac{22.488}{4.157} = 5.410$$

The tabulated value of the Fisher-Snedecor test for a significance level of α= 0.05 and degrees of freedom $F_{0.95}\,(N - l, n_0 - 1) = F_{0.95}(18,\ 2) = 19.4$. Since this tabulated value is greater than the calculated value, the model is therefore considered unbiased.

- ***Regression Significance Test***

Pour le niveau de signification α= 0.05 et les nombres de degré de liberté (N- l)= 18 et (l -1)= 8. la valeur tabulée du teste de Fisher est de 2.51. La valeur de F (37.632) étant supérieure à celle tabulée. l'équation de régression est adéquate et le modèle est valide à 95 %. L'équation retenue pour le modèle s'écrit donc : For a significance level of α= 0.05 and degrees of freedom (N- l)= 18 and (l -1)= 8, the tabulated value of the Fisher test is 2.51. Since the calculated value of F=37.632 is greater than the tabulated value, the regression equation is adequate and the model

is valid at the 95% confidence level. The retained equation for the model is therefore written as:

$$\hat{y} = 71.287 - 3.437\,x_1 - 11.490\,x_2 + 5.503\,x_3 + 6.395\,x_4 + 2.973\,x_2x_4 - 2.012\,x_1^2 + 4.416\,x_3^2 + 4.709\,x_4^2$$

- ***Model validation using the center point***

The degradation yield predicted by the model at the center of the study domain is $\hat{y}$ = 71.287 % which is equal to b_0 and the value obtained from the three center-point runs is $\overline{y}_0$= 71.287 %. This confirms that the obtained model is adequate and accurately simulates the experiment.

- ***Calculation of the Coefficient of Determination***

The value of the coefficient of determination is:

$$R^2 = \frac{\sum_{i=1}^{N}(\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} = 0.943$$

And that of the adjusted coefficient of determination is:

$$\overline{R}^2 = R^2 - (1 - R^2)\frac{l-1}{N-l} = 0.920$$

These values allow us to assert that the chosen model is adequate.

- **Residual Analysis**

The quality of the second-order model can be evaluated through a residual analysis (Fig. IV.4).

Calculation of the Residuals Mean: $Moy_{résiduals} = \frac{\sum e_i}{N}$ x 100

Where e_i is the residual and N is the number of runs.

$$Moy_{résiduals} = 4.533 \times 10^{-14}$$

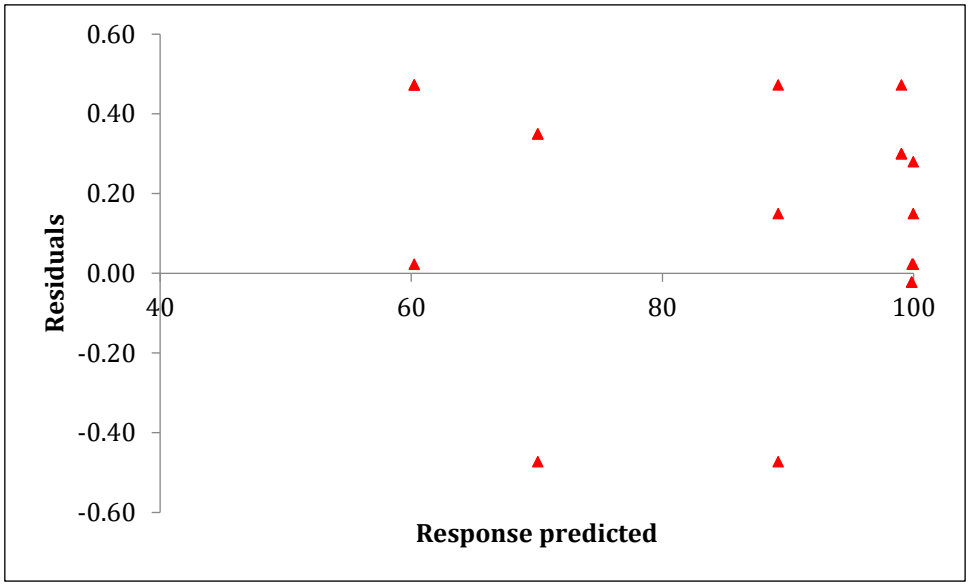


Figure IV.4: Residual Plot

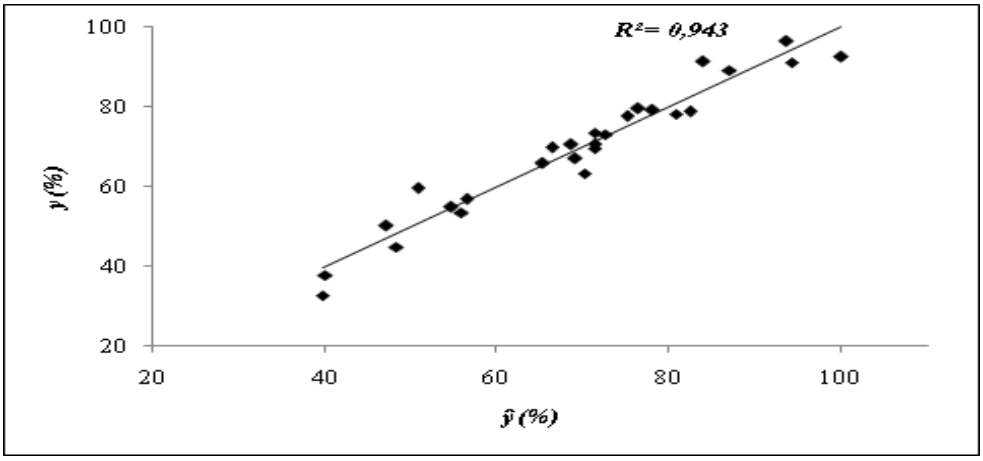


Figure IV.5. Comparison between the Predicted Response Values and the Experimental Values

➢ Optimization

Based on the second-order model developed, the search for the optimal point can be performed:

- By solving the model equation;
- By plotting the response surfaces and iso-response curves.

❖ Solving the Model Equation

From the equation below, we can calculate the optimal values of the operating parameters that lead to the maximum dye degradation yield. To do this, it is sufficient

to solve the system of equations obtained by differentiating the predictive variable with respect to each of the parameters x_1, x_2, x_3 and x_4.

$$\hat{y} = 71.287 - 3.437\,x_1 - 11.490\,x_2 + 5.503\,x_3 + 6.395\,x_4 + 2.973\,x_2x_4 - 2.012\,x_1^2 + 4.416\,x_3^2 + 4.709\,x_4^2$$

$$\begin{cases} \dfrac{\delta\hat{y}}{x_1} = 0 \Leftrightarrow -3.437 - 4.024\,x_1 = 0 \\ \dfrac{\delta\hat{y}}{x_2} = 0 \Leftrightarrow -11.49 + 2.973\,x_4 = 0 \\ \dfrac{\delta\hat{y}}{x_3} = 0 \Leftrightarrow 5.503 + 8.832\,x_3 = 0 \\ \dfrac{\delta\hat{y}}{x_4} = 0 \Leftrightarrow 6.395 + 2.973\,x_2 - 9.418\,x_4 = 0 \end{cases}$$

Solving this system of equations leads to:

$$\begin{cases} x_1 = -0.854 \\ x_2 = 10.093 \\ x_3 = -0.623 \\ x_4 = 3.865 \end{cases}$$

The values of x_2 and x_4 fall outside the study domain. We will set them to their maximum values (+2). The optimal values of the different parameters are:

Q_v = 159.300 mL /min, $[JV]_0$= 50 mg/L, pH= 5.410 and $[H_2O_2]$ = 400 mg/L.

The theoretical yield obtained at this point (- 0.854, +2, -0.623, +2) is 25.42%. This optimum corresponds to a minimum.

❖ ***Response Surfaces and Iso-Response Curves***

Figures IV.6 and IV.7 show the response surfaces and iso-response curves plotted using the STATISTICA software in the x_2 and x_3 plane (initial dye concentration – pH), while fixing the variables Qv= 159.3 mL/min and $[H_2O_2]_0$= 400 mg/L. Their shapes indicate that a maximum yield exists at the top of the surface. This yield can be achieved with different combinations of pH and $[JV]_0$ within the domain: 9.125-10 for pH and 5-26.34 for $[JV]_0$ (Figure IV.7).

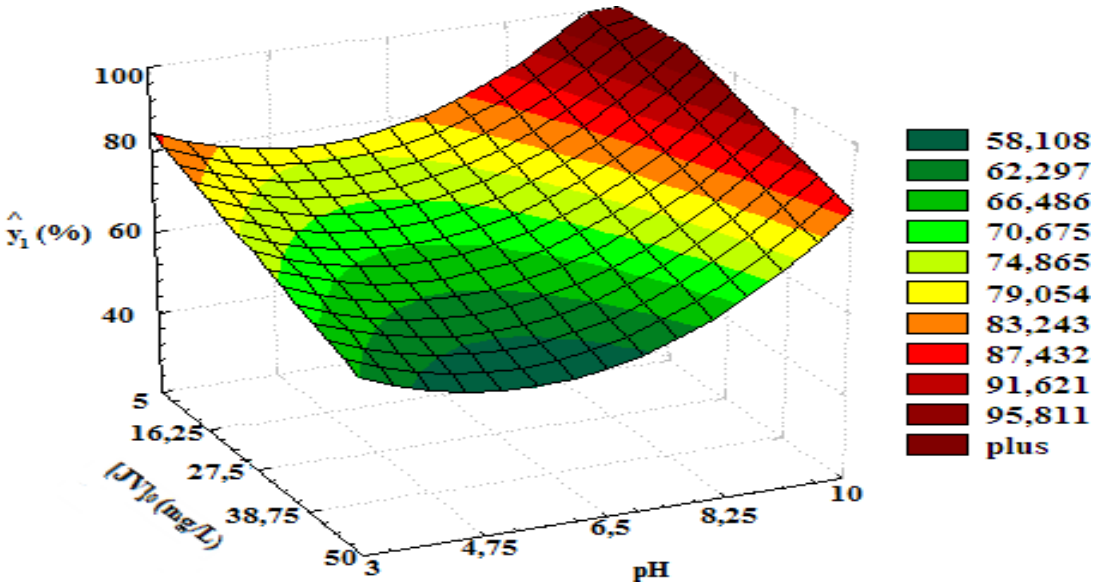


Figure IV.6 Response surface in the pH–initial concentration plane for the dye degradation rate (Q_v = 159.3 mL/min and [H_2O_2] = 400 mg/L).

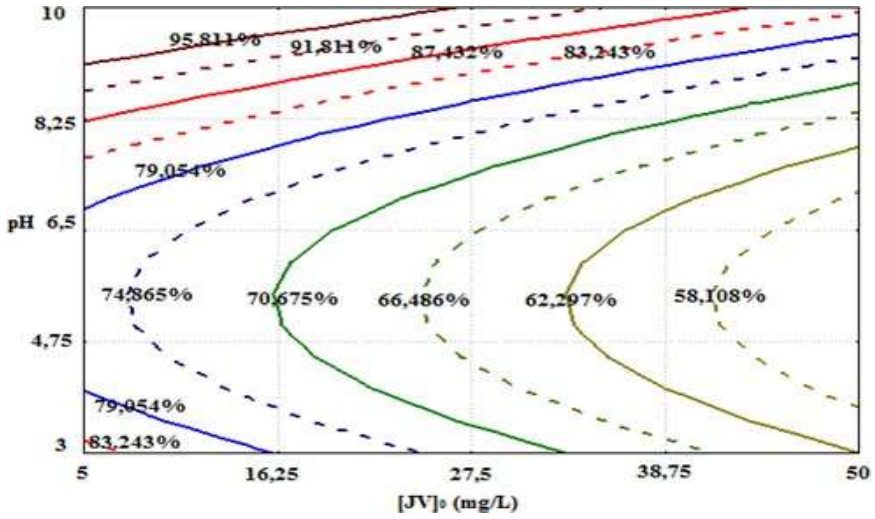


Figure IV.7. Iso-response curves in the pH–initial concentration plane for the dye degradation rate (Q_v = 159.3mL/min and[H_2O_2] = 400mg/L).

Chapter V
Mixture designs

V.1 Introduction

In chemical engineering, the study of mixtures is essential for the design and optimization of many processes, whether related to product formulation, chemical reactions, or phase separations. Unlike traditional experimental designs, where factors can vary independently, mixture designs take into account a fundamental constraint: the sum of the mass, molar, or volumetric fractions of the components must remain constant, generally equal to 1. This specific characteristic requires a particular experimental approach in order to properly model the properties of the mixture and predict its behavior.

The use of mixture designs thus makes it possible to analyze the influence of each component, identify potential interactions, and optimize formulations and processes with the aim of improving the overall performance of the systems studied in chemical engineering.

V.2 Geometrical Representation of Mixtures

V.2.1 Two-Component Mixture

Let x_1 be the proportion of the first component and x_2 that of the second. Let us adopt a Cartesian representation, where the Ox_1 axis is orthogonal to the Ox_2. The axes are graduated in proportions ranging from 0 to 1. Any mixture containing x_a of A and x_b of B is represented by a point located at the intersection of the coordinates x_a and x_b (Fig. V.1). This point, which symbolizes a mixture, is referred to as a composition point, mixture point, or simply a point.

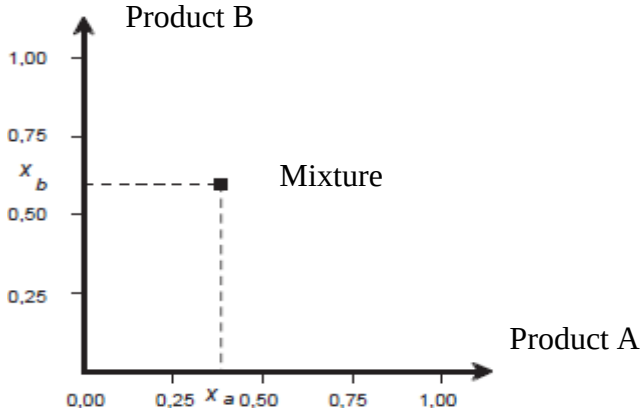


Figure V.1: Representation of a mixture in a cartesian coordinate system

The mixture constraint introduces a relationship between x_a and x_b :

$$x_a + x_b = 1 \text{ which can be written as: } x_b = -x_a + 1$$

This relationship indicates that the points with coordinates x_a and x_b lie on a straight line with a slope of (−1). This line intersects the axis corresponding to product A at the abscissa 1.00 (point A in Figure V.2) and the axis corresponding to product B at the ordinate 1.00 (point B in Figure V.2).

Since the compositions x_a and x_b vary between 0 and 1, only the line segment AB is meaningful within the experimental domain. All possible compositions of mixtures composed of products A and B are therefore represented by the points located along this line segment.

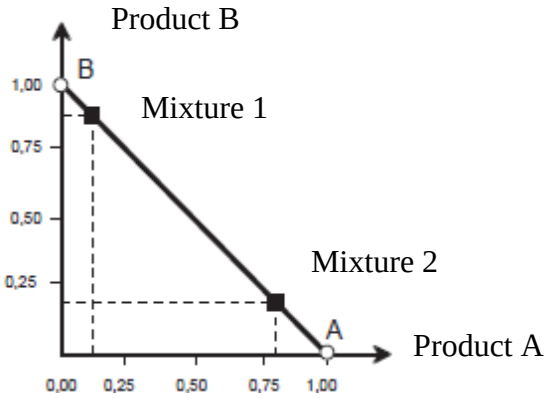


Figure V.2: Representation of the composition domain of a binary mixture in the (x_A, x_B) plane

Therefore, only this line segment can be retained, and the Ox_1 and Ox_2 axes may be omitted. Pure component A is represented by one end of segment AB, while pure component B is represented by the other end. This segment is provided with a double scale (Fig. V.3): one corresponding to the fraction of the first component and the other to the fraction of the second component. Reading the mixture composition along this segment requires some practice, as it must be interpreted both from right to left and from left to right.

V.2.1.1 Reading of the binary mixture diagram

The segment shown in Figure V.3 represents the mixtures of the two components, A and B. Component A is located on the right, while component B is on the left. The lower scale corresponds to component A; it ranges from 0 on the left to 1 on the right. The upper scale corresponds to component B; it ranges from 0 on the right to 1 on the left.

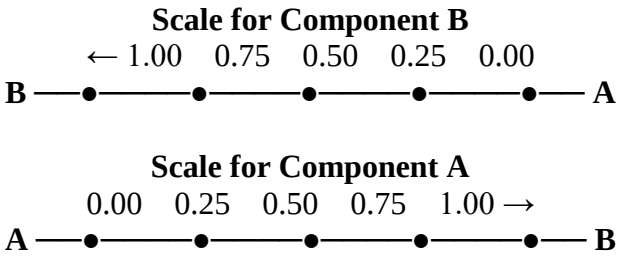


Figure V.3: Representation of binary mixtures on a straight line segment

The scales are read in opposite directions. At any point on the segment, the sum of the component fractions equals unity.

V.2.2 Three component mixture

To represent a mixture of three components (A, B, and C), an equilateral triangle is used (Fig. V.4).

- Each vertex of the triangle corresponds to a pure component.
- Each side represents a binary mixture of two components, excluding the third.
- The sides are graduated from 0 to 1 for one component and the proportion of the other is obtained by subtracting from 1.

A point inside the triangle represents a mixture of all three components. To determine the proportion of each component:

- The point is projected onto the corresponding side along a direction parallel to the opposite side.

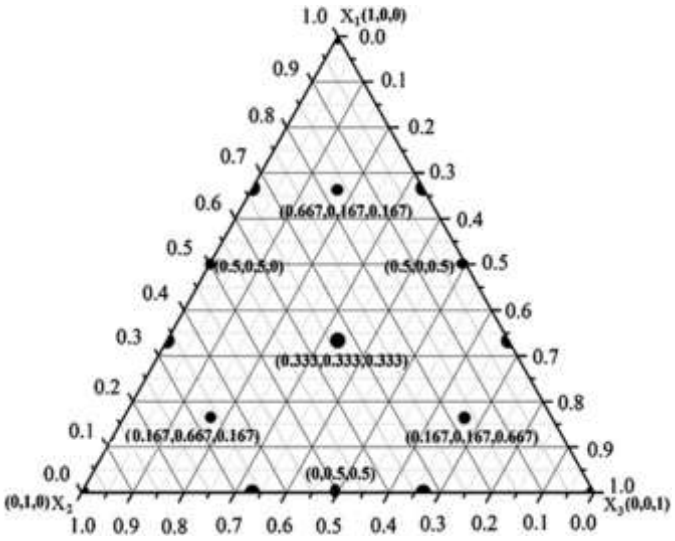


Figure V.4: Graphical representation of the composition of a three-component mixture using a ternary diagram.

V.2.2.1 Classical mixture designs

Classical mixture designs assume that the pure components possess the desired functional properties and that no constraints are imposed on them. The purpose is to classify the main types of mixture designs according to the location of the points representing the compositions. The main types are:

- Lattice designs,
- Centered mixture designs,
- Augmented centered mixture designs.

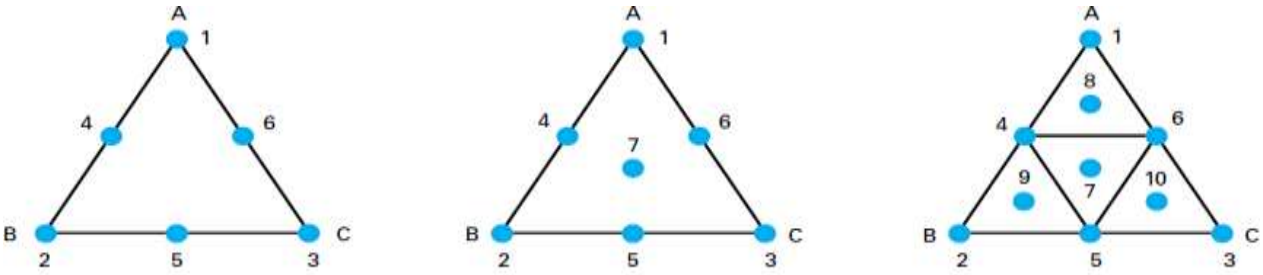


Figure V.5: Lattice mixture design (left), centered mixture design (middle), augmented centered mixture design (right)

V.2.3 Four component mixtures

To represent a mixture of four components, a regular tetrahedron is used (Fig. V.6).

- Each vertex corresponds to a pure component.
- Each edge represents a binary mixture of two components.
- Each face (triangle) represents a ternary mixture of three components.

The interior of the tetrahedron represents mixtures of all four components. To determine the composition, the mixture point is projected onto the faces and edges of the tetrahedron.

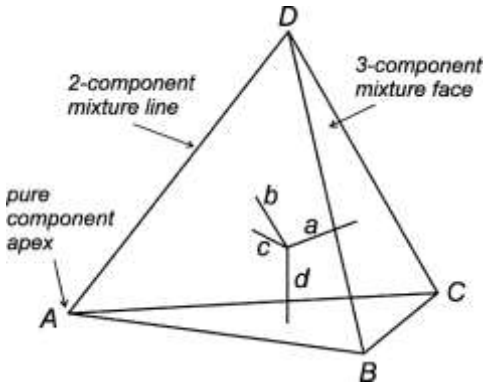


Figure V.6: Graphical Representation of Four-Component Mixtures by Means of a Regular Tetrahedron

V.3 Mathematical models for mixture designs

As in classical experimental design, the variation of the response y^\hat{y}y^ as a function of mixture composition is generally described using standard polynomial models. By incorporating the constraint that the sum of the component fractions equals unity, these models can be reformulated into a simplified expression known as the canonical form.

V.3.1 First-order model

For a three-component mixture, the proposed first-order model is obtained from the classical first-degree polynomial expressed as:

$$\hat{y} = a_0 + a_1x_1 + a_2x_2 + a_3x_3 \qquad (Eq.\,V.1)$$

To this expression, we apply the fundamental mixture constraint:

$$x_1 + x_2 + x_3 = 1$$

Where:

$\hat{y}$: predicted response estimated by the model for a given mixture composition.

Equation (V.1) can then be rewritten as follows:

$$\hat{y} = a_0(x_1 + x_2 + x_3) + a_1x_1 + a_2x_2 + a_3x_3 \qquad (Eq.\,V.2)$$

By regrouping the coefficients, the model takes the following form:

$$\hat{y} = (a_0 + a_1)_1 + (a_0 + a_2)_2 + (a_0 + a_3)x_3$$

Hence, the model can be expressed as:

$$\hat{y} = b_1x_1 + b_2x_2 + b_3x_3 \qquad (Eq.\,V.3)$$

$$\text{with: :}\begin{cases} b_1 = a_0 + a_1 \\ b_2 = a_0 + a_2 \\ b_3 = a_0 + a_3 \end{cases}$$

Where $b_1, b_2\ and\ b_3$ are the linear mixture coefficients of the model.

V.3.3 Second-Order Model

The second-order model including interaction (synergistic) terms is expressed as follows:

$$\hat{y} = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{23}x_2x_3 + a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2$$
(Eq. V.4)

By applying the fundamental mixture constraint:

$$\hat{y} = b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 \quad \text{(Eq.V.5)}$$

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

and

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

V.3.4 Third-Order Models

The third-degree mathematical model, specific to mixture designs, is obtained by applying the fundamental mixture constraint to a third-degree polynomial. It is expressed in the following form:

$$\hat{y} = b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 \quad (Eq.V.6)$$

This simplified model is referred to as the reduced cubic model. It includes only the linear terms as well as the binary and ternary interaction terms.

V.4 Application Example : *Adsorption of pharmaceutical residues on adsorbents prepared from olive stones using mixture design of experiments model. Water Science & Technology 80 (5): 998–1009, 2019.*

This study aims to investigate the relationship between the amount of tetracycline (TC) adsorbed (y) and three adsorbent materials. unmodified olive pomace (UOS: Sorbent 1), $ZnCl_2$-activated carbon derived from olive pomace (ACOS $ZnCl_2$: Sorbent 2), and H_3PO_4-activated carbon derived from olive pomace (ACOS H3PO4: Sorbent 3).

$$y = q_t = \frac{(C_0 - C_t)}{m} V$$

Where:

C_0 : initial concentration of tetracycline (mg/L);

C_t : concentration of tetracycline in the solution at time t (mg/L);

m : mass of the adsorbent (g);

V: volume of the solution (L).

The selected design is an augmented simplex centroid mixture design, which includes the mixtures listed in Table V.1. The experimenter plans to use points 8, 9 and 10 as control points (Fig. V.7). These points will not be used to develop the mathematical model but will serve to verify the predictive capability of the model (Table V.2).

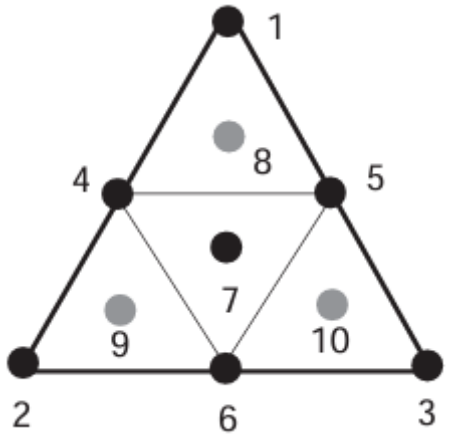


Figure V.7 Position of the experimental points (black) and control points (gray) in the applied design

TableV.1: Matrice de plan d'expériences

N° \ Sorbents	Sorbent 1	Sorbent 2	Sorbent 3	y = q_t (mg/g)
1	1	0	0	39.64
2	0	1	0	42.01
3	0	0	1	97.59
4	0.5	0.5	0	43.59
5	0.5	0	0.5	56.49
6	0	0.5	0.5	77.30
7	0.3333	0.3333	0.3333	58.87

✓ **Estimation of the model coefficients**

The coefficients of the regression equation are estimated using the following expression:

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

✓ **Elaboration d'un modèle de second degré**

Development of a second-degree model:

$$\hat{y} = 39{,}64\ x_1 + 42{,}01\ x_2 + 97{,}59\ x_3 + 11{,}46 x_1 x_2 - 48{,}46 x_1 x_3 + 30{,}02 x_2 x_3$$

✓ **Validation of the second-degree model using control points**

The results comparing the experimentally obtained responses with those predicted by the model are presented in Table V.2.

Table V.2 : Validation of the models using control points

Sorbents / N°	Sorbent 1	Sorbent 2	Sorbent 3	y (Experimental)	y (Predict)	Erreurs (%)
7	0.3333	0.3333	0.3333	58.87	58.61	0.44
8	0.6666	0.1666	0.1666	47.48	45.70	3.74
9	0.1666	0.6666	0.1666	54.09	51.67	4.68
10	0.1666	0.1666	0.6666	76.87	78.09	1.56

The same statistical analysis described in Chapter II was performed to validate the proposed mathematical model prior to its interpretation.

Figure V.8 presents the response surface (a) and the iso-response contours (b), highlighting the influence of the proportions of Sorbent 1 (x_1), Sorbent 2 (x_2), and Sorbent 3 (x_3) on the adsorption capacity of TC.

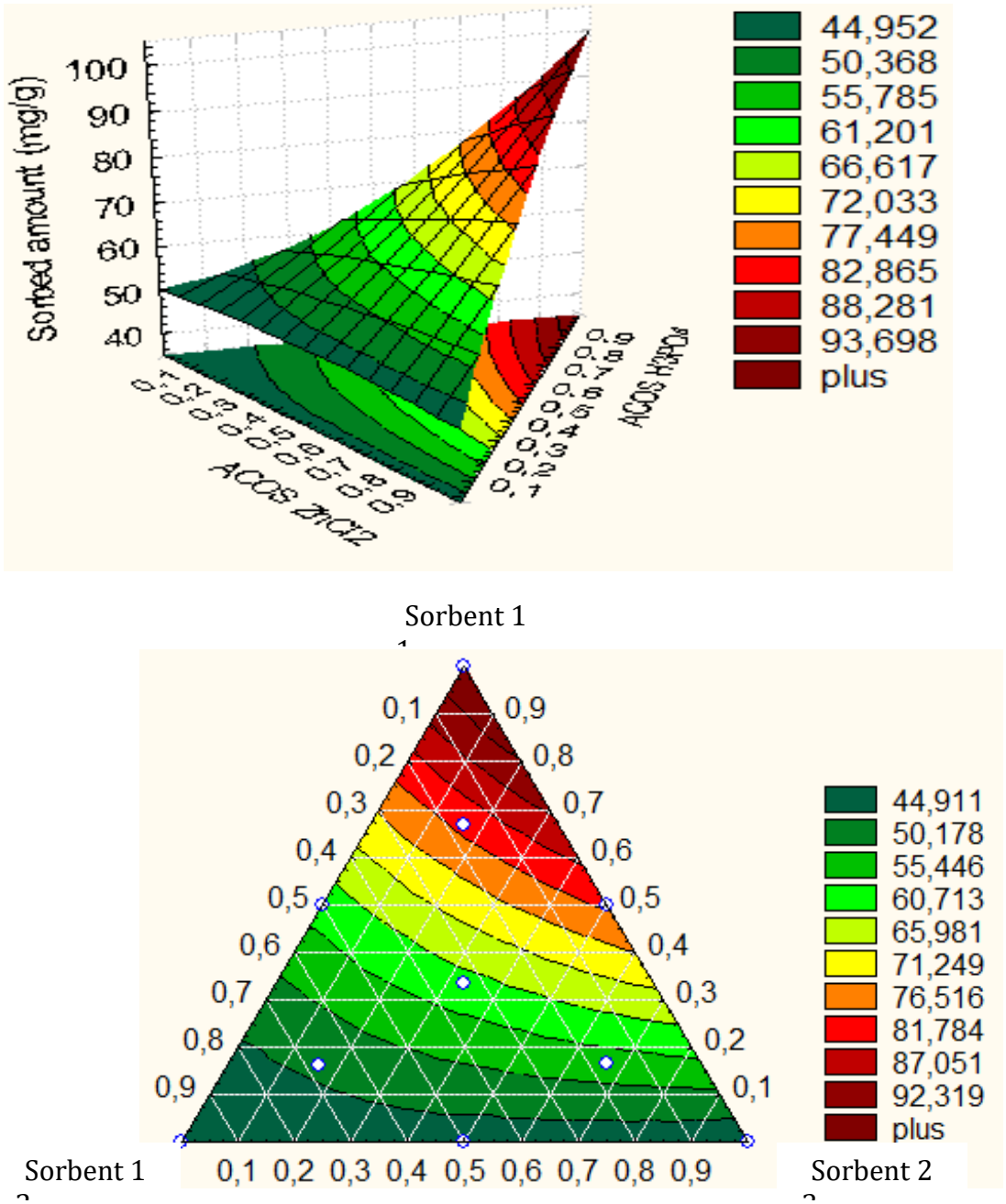


Figure V.8. Response surface plots (a) and corresponding iso-response contours (b) illustrating the influence of the proportions of Sorben 1 (x_1), Sorben 2 (x_2), and Sorben3 (x_3) on the adsorption capacity of TC.

V.6 Mixture designs and experimental designs: combined (mixture–process) designs

To simultaneously investigate the effects of mixture component proportions and other experimental factors, it is necessary to combine a mixture design with a conventional experimental design (Fig. V.9).

For example, in the case of a chocolate manufacturer:

•The chocolate formulation is studied using a mixture design.
• The processing conditions (such as temperature or baking time) are investigated using a factorial or response surface design.

For each experimental condition defined by the experimental design, a complete mixture design must be performed. Consequently, the total number of experiments increases rapidly. If n runs are required for the mixture design and p runs for the experimental design, the total number of experiments is equal to $n \times p$.

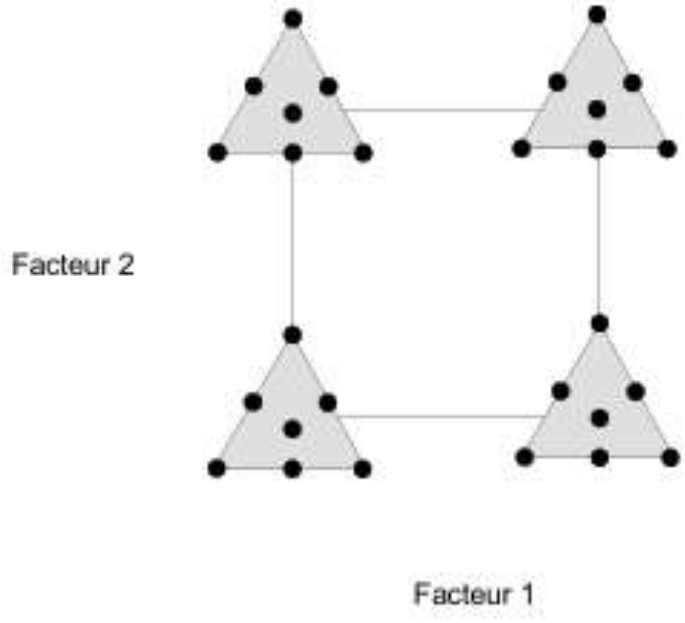


Figure V.9: Graphical representation of a three-component mixture study under the influence of two process factors.

V.7 Software for design of experiments

Experimental design software not only provides libraries of standard designs but also allows the creation of customized plans. In particular, the construction of mixture designs and designs with specific constraints on the experimental domain requires the use of software to generate the plan best suited to the study (Table V.3)

Table V.3: Main software programs for design of experiments

JMP	*http://www.jmp.com*
MINITAB	*http://www.minitab.fr*
Statistica	*http://www.intesoft.com/produits/tech/statistica*
Statgraphics	*http://www.sigmaplus.fr*
Unscrambler	*http://www.camo.com*
Pirouette	*http://www.infometrix.com*
Modde	*http://www.umetrics.com*

Conclusion

By following this course, students will gain the knowledge and practical skills required to design, analyze, and interpret experiments efficiently. Emphasis is placed on applying theoretical principles to real-world problems, enabling students to optimize processes, make informed decisions, and contribute effectively to research and industrial projects. Mastery of these experimental design techniques will therefore provide a solid foundation for both academic and professional success in the fields of engineering and applied sciences.

Bibliographic references

- F. Boudrahem, I. Yahiaoui, S. Saidi, K. Yahiaoui, L. Kaabache, M. Zennache and F. Aissani-Benissad. Adsorption of pharmaceutical residues on adsorbents prepared from olive stones using mixture design of experiments model. Water Science & Technology 80 (5): 998–1009, 2019.
- I. Yahiaoui and F. Aissani-Benissad. Experimental design for copper cementation process in fixed bed reactor using two level factorial design. Arabian Journal of Chemistry 3: 187, 2010.
- I. Yahiaoui, A. Belattaf, F. Aissani-Benissad and L. Yahia-Cherif. Full factorial design applied to a biosorption of lead (II) ions from aqueous solution using a brewer's yeast (Saccharomyces cerevisiae). Journal of Chemical & Engineering Data 56: 3999, 2011.
- I. Yahiaoui, F. Aissani-Benissad and 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, 2010.
- I. Yahiaoui. Les plans d'expériences Initiation à la construction et l'analyse des plans factoriels complets à deux niveaux et du plan composite centré (Course notes). Université A. MIRA, Bejaia, 2015.
- J. Goupy, Introduction aux plans d'expériences avec application, Edition DUNOD, Paris, 2013.
- J. Goupy, Introduction aux plans d'expériences, Edition DUNOD, Paris, 2001.
- J. Goupy, Les plans d'expériences : Les mélanges, Edition DUNOD, Paris, 2000.
- J. Goupy, Modélisation par les plans d'expériences, Techniques de l'Ingénieur, R275, 2000.
- J. Goupy, Plans d'expériences pour surfaces de réponse, Edition DUNOD, Paris, 1999.
- J. Goupy, Plans d'expériences, Techniques de l'Ingénieur, PE 230, 1997.
- J. Goupy, Pratiquer les plans d'expériences, Edition DUNOD, Paris, 2005.
- L. Yahia Cherif, I. Yahiaoui, F. Aissani-Benissad, K. Madi, N. Benmehdi, F. Fourcade, and A. Amrane, Heat attachment method for the immobilization of TiO_2 on glass plates: Application to photodegradation of basic yellow dye and optimization of operating parameters, using response surface methodology. Industrial & Engineering Chemistry Research 53, 10, 3813–3819, 2014.
- P. Ozil, Plans d'expériences : Méthode de TAGUCHI, Ecole Nationale Supérieure

d'Electrochimie et d'Electrométallurgie, INPG, 1997.

- S. Ziani. Les plans d'expériences (Course notes). Université Akli Mohand Oulhadj, Bouira. 2019.
- Z. R. Lazic, Design of Experiments in Chemical Engineering, Edition WILEY-VCH, Weinheim, 2004.