

UNIVERSITE DE LIMOGES

ECOLE DOCTORALE Science – Technologie – Santé

FACULTE de des Sciences et Techniques

Année : 2004

Thèse N° 39-2004

Thèse

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE DE LIMOGES

Discipline / Spécialité : Electronique des hautes fréquences et optoélectroniques

Présentée et soutenue par

Sandrine KARAM

Le 26 novembre 2004

Application de la méthodologie des plans d'expériences
et de l'analyse de données à l'optimisation des
processus de dépôt

Thèse dirigée par Annie Bessaudou

JURY :

André LEBLE	Université du Mans	Rapporteur
Jean-Paul RIVIERE	Université de Poitiers	Rapporteur
Annie BESSAUDOU	Université de Limoges	Examinatrice
Claire DARRAUD	Université de Limoges	Examinatrice
Olivier TANTOT	Université de Limoges	Examinateur
Pascal TRISTANT	Université de Limoges	Examinateur
François LOUVET	E.N.S.C.I. de Limoges	Invité

A ma famille, mes amis et à l'ouverture d'esprit...

Résumé

Traditionnellement, un chercheur effectue des expériences de manière séquentielle en faisant varier les paramètres les uns après les autres. Cette méthode donne des résultats mais elle est coûteuse en temps et nécessite la réalisation d'un grand nombre d'expériences. Les méthodes d'analyses de données permettent de collecter, de résumer et de présenter des données de façon à en tirer le maximum d'informations pour la suite des expériences. Pour mener une recherche expérimentale planifiée, la méthodologie des plans d'expériences est utilisée. Cette démarche aide l'expérimentateur à structurer sa recherche de manière différente, à valider ses propres hypothèses, à mieux comprendre les phénomènes étudiés et à solutionner les problèmes. L'objectif de cette thèse est de présenter, à travers des exemples d'applications liés à l'élaboration de dépôts, la mise en œuvre et l'intérêt de ces méthodes destinées à tout expérimentateur désireux d'optimiser l'efficacité de sa recherche expérimentale.

Mots clés :

- Plans d'expériences
- Criblage
- Plan à surface de réponses
- Analyse de données
- Dépôts
- Couches minces

Abstract

Experimental designs and data analyses for the optimization of deposition processes.

Traditionally, a scientist carries out experiments in a sequential way by varying the parameters the ones after the others. This method gives results but is very expensive in time because it inevitably requires the realization of a great number of experiments. This is why it is important to help the scientist to achieve his experiments with data analyses and experimental design methods. Data analysis makes it possible to collect, summarize and present data so as to optimize the way to implement next experiments. By using experimental design, the scientist knows how to plan experiments. This experimental step will help him to structure his research in a different way, to validate his own assumptions, with better understanding the studied phenomena, and to solve the problems. The objective of this thesis is to present, through examples of applications related to the development of deposits, the elaboration and the interest of these methods.

Keywords :

Experimental design

Screening

Responses surfaces

Data analyses

Deposition processes

Thin films

Remerciements

Cette étude a été réalisée à la Faculté des Sciences et Techniques de l'université de Limoges au sein de l'Institut de Recherche en Communications Optiques et Microondes (IRCOM).

Je remercie avec une amitié sincère ma directrice de thèse Annie BESSAUDOU avec qui j'ai tissé au fil des jours une relation franche et très agréable qui m'a toujours soutenue et comprise. Je remercie également son mari Eric (au « mille conversations passionnantes ») qui m'a toujours amenée dans ses bagages pour les fameuses journées « Expérimentiques » (ou « andouillette frites »). Je remercie également toute l'équipe C2M avec qui il fait bon fêter les anniversaires.

J'exprime également mes profonds remerciements à Monsieur André LEBLE et Monsieur Jean-Paul RIVIERE, qui, fort de leurs connaissances et de leurs expériences, ont eu à juger ce travail.

Toute ma gratitude revient à Claire DARRAUD, Olivier TANTOT, Pascal TRISTANT qui ont accepté de faire partie du jury.

J'attire une grande attention sur un grand génie de l'analyse de données et des plans d'expériences « tout fait main », François Louvet pour ses explications, ses écrits, sa pédagogie et sa philosophie sans qui je n'aurais pu soutenir cette thèse.

Je remercie chaleureusement et tendrement mes parents qui m'ont soutenue et poussée à faire des études, parfois aux prix de quelques sacrifices et sans qui je n'aurais pas pu atteindre ce niveau.

Une pensée toute particulière à une Grande Belle Blonde Intelligente et parfois Machiavélique nommée Mimi que j'aime énormément et qui me soutient jour après jour avec qui j'entretiens une relation pleine de complicité.

Je remercie tous mes collègues et à présent amis compagnons de thèse : ma compagne de breuvage de « Ste » dite Stéphanie (confidente, rigolote et en plus actrice de road movie) pour sa simple douce et réconfortante présence à mes cotés durant ces années et plus encore, Cyrio dit le Padre (grand, très grand acteur !) pour « on est pas bien là ! », Guéno (frisouille tendre et sensible) pour son soutien et sa disponibilité à toutes épreuves, Ben (hum c'est bon de manger !) pour nos discussions psychologiques, Win's (monteur et blagueur né) pour son humour débordant, la Parvs (dit Alec Baldwin surtout l'été) pour sa mèche et sa purée mousseline, ma cops Sandrine (dite la grande) et ses parents Jeannot et Jeannette que j'adore qui me fêtent le Noël en février, Bob pour

ses bateaux de moules à roue, Geo (cinéophile rigolard) pour son Hi hi hi, GG (relookage de « c mon choix »), Val (bijout girl) et Laure (dite la plus petite de nous deux) pour les soirées gonzesses, Vic (batteur de jazz) pour mes années facs thé discussions, Toto pour son humeur bacrisante et sa collaboration, Tibo et Airv (dits « les gars du spcts ») pour notre recherche commune et leurs croyances en ces méthodes, à mes collègues d'équipe Marc (dit peigne cul), Pierre, Manu, Sylvain (qui fait de belles photos maintenant !) et tout particulièrement Matthieu sans qui les débats n'ont pas d'intérêts.

J'attire une attention particulière sur un personnage sorti tout droit de la science fiction dit Aimé Leblanc, grand adolescent à la tignasse grisonnante mais au cœur d'un jeune de 20 ans, ses blagues, son bureau fumoir, son journal et son attention permanente en fait de lui le personnage clé du laboratoire, c'est notre guide à tous, merci encore Mémé... on t'aime. Un grand merci à Armand et Benoît (dit Lolo) sans qui ma thèse ne serait pas reliée aujourd'hui ainsi qu'à Françoise (la step girl au petit sac rose bonbon) et cathy.

Je remercie ma nouvelle famille, mes colocataires d'un certain manoir, la Maude (belle humanitaire au cœur chaud), la hyène (Inspecteur de frère et sœur), Yann (Humoriste radical), Fred (Punck de velours), Ptit bibi (hihihihi) et Guéri (le cafard fitoxé) mes copains de la cour de récré quotidienne, Marioun's (Adorable douce Stretching girl), Creusois et Tête d'oïse (Humains aux ailes synthétiques, « Et Tac le Win ») qui m'ont chaperonnée durant ma rédaction, ma seconde soeur Ekila (« Yes I ») que j'aime appeler pour rien dire et que je vois vieillir depuis des années en même temps que moi (c'est rassurant de voir les gens qu'on aime nous accompagner!), Cathy (working girls parisienne) pour ses conseils terribles et avisés d'une pro, ma petite Valérie jumelle, la ptite manue et salomé, laeti, Nanouk et Mat qui m'ont hébergée, mon fidèle parapentiste de pote Vincent, mon ami Tidams qui croit à mes cotés créatifs, ma collègue de thèse Aurélie et son futur « fiancé » dit le Boutch, Isa, JT et Léon dit la famille Gaston la Gaffe. Je tiens également à citer des personnages clé de ce tiers de vie passée dans cette faculté : mon pote Momo (guyanais d'un temps, hum c bon le rhum arrangé), pierre (star des ondes du Move) qui est un des rares à avoir réaliser ses projets professionnels de départ, mon amie Cynthia avec qui j'ai partagé de longues discussions sur lesquelles j'ai bâti ma personnalité, Sabrina (dite « short P... ») que j'apprécie énormément pour ces conseils esthétiques, Cyril (dit « la mère ») avec qui j'ai raté mes concours d'école de ciné, le Gais (avec plaisir), Jugurtha (la crème), Seb (la théorie du gant de toilette), Jean-charles (rage against the machin informaticien), mon ami éternellement dans mon cœur julien, ma douce Christounip et ma futur morue.

Une pensée à guillaume et à sa famille...

SOMMAIRE

INTRODUCTION GENERALE 1

PARTIE A : RAPPELS THEORIQUES SUR LES PLANS D'EXPERIENCES

Introduction 5

Chapitre I : Origine et état de l'art..... 6

I.1. Pourquoi faire des plans d'expériences ?..... 6

I.1.1. La recherche exploratoire 6

I.1.2. Méthode traditionnelle des essais – erreurs 6

I.2. Méthodologie des plans d'expériences..... 8

I.2.1. Qu'est ce qu'un plan d'expériences ? 8

I.2.2. Historique et état de l'art 8

Chapitre II : Formalisation du problème..... 10

II.1. Introduction..... 10

II.2. Définir le problème..... 10

II.3. Déterminer les objectifs..... 11

II.4. Définir les sorties (réponses)..... 11

II.5. Définir les entrées (facteurs)..... 12

II.5.1. Phase de créativité 12

II.5.2. Phase de classement 12

II.5.3. Définition des facteurs 14

II.5.4. Domaine expérimental 15

Chapitre III : Choix d'une stratégie..... 17

III.1. L'efficacité des plans d'expériences..... 17

III.1.1. Stratégie traditionnelle : « un seul facteur à la fois » 18

III.1.2. Stratégie exhaustive : le plan factoriel complet 18

III.1.3. La stratégie du plan fractionnaire orthogonal 19

III.2. Construction du plan d'expérimentation..... 22

III.2.1. Choix de la stratégie expérimentale 22

III.2.2. Choix du modèle 22

III.2.3. Choix de la matrice d'expériences 23

III.2.4. Les contraintes particulières à la construction d'un plan 24

III.2.5. Définir le domaine expérimental 24

Chapitre IV : Analyses et résultats..... 25

IV.1. Acquisition progressive des résultats..... 25

IV.2. Analyse globale des résultats d'essais..... 25

IV.3. Analyse mathématique..... 26

IV.3.1. Notations mathématiques 26

IV.3.2. Matrice du modèle, matrice d'information et de dispersion 26

IV.3.3. La régression multilinéaire 28

IV.4. Analyse statistique des résultats..... 29

IV.5. Analyse graphique des résultats..... 34

IV.6. Validation des hypothèses et recherche de solutions au problème..... 34

Conclusion..... 36

Rappel des notations..... 37

PARTIE B: THEORIE SUR L'ANALYSE DE DONNEES

Introduction	40
I.1. Etat de l'art.....	40
I.2. Historique et définition.....	40
I.3. Utilisation de l'analyse de données.....	40
.	
Chapitre II Pré-traitement des données.....	42
II.1. La nature des données	42
II.1.1. Définition des variables	42
II.1.2. Définition des individus	42
II.1.3. Notations	42
II.1.4. Codage du tableau de données brutes	44
II.2. Statistique élémentaire.....	46
II.2.1. Analyse unidimensionnelle de données	46
II.2.2. Analyse bidimensionnelle	50
II.3. Organigramme	55
Chapitre III L'analyse en composante principale (A.C.P.).....	56
III.1. Introduction.....	56
III.2. Présentation générale sur l'A.C.P.....	56
III.2.1. Introduction	56
III.2.2. Quels types de problèmes peuvent être abordés par l'a.c.p. ?	56
III.2.3. Quels tableaux de données peut-on traiter avec l'a.c.p. ?	57
III.2.4. Principe de l'a.c.p.	57
III.3. Démarche d'une analyse factorielle.....	58
III.3.1. Construction des axes principaux, démarche géométrique	58
III.3.2. Notion d'inertie : démarche analytique	59
III.4. Présentation de l'A.C.P. à partir d'un exemple.....	61
III.4.1. Bilan des analyses précédentes	61
III.4.2. Part d'information portée par les axes	61
III.4.3. Nouvelles coordonnées des individus et des variables	63
III.4.4. Contribution des individus au calcul des axes	64
III.4.5. Qualité de représentation des individus	65
III.4.6. Outils graphiques : le cercle des corrélations et le biplot	67
Chapitre IV La classification ascendante hiérarchique (C.A.H.).....	72
IV.1. Introduction.....	72
IV.2. Principe.....	72
IV.3. Arbre de classification.....	76
Chapitre V L'Analyse Factorielle Discriminante (A.F.D.).....	78
V.1. Introduction.....	78
V.2. Principe.....	78
Conclusion sur l'analyse de données.....	84
Rappel des notations.....	85

PARTIE C: EXEMPLE D'APPLICATIONS DE L'ANALYSE DE DONNEES ET DES PLANS D'EXPERIENCES

Chapitre I : Optimisation des couches minces d'alumine réalisées par PECVD

I.1. Formalisation du problème.....	87
I.1.2. Variables de sortie	89
I.1.3. Variables d'entrée	91
I.2. Criblage des facteurs.....	91
I.2.1. Facteurs – modalités – domaine expérimental	91
I.2.2. Modèle empirique	93
I.2.3. Construction de la matrice du modèle	94
I.2.4. Expérimentations	98
I.2.5. Analyse globale des résultats d'essai	98
I.2.6. Analyse mathématique des résultats d'essai	100
I.2.7. Analyse graphique du modèle	105
I.2.8. Analyse statistique du modèle	107
I.2.9. Bilan et conclusion	109
I.3. Optimisation – plans a surfaces de réponses.....	110
I.3.1. Introduction	110
I.3.2. Facteurs et niveaux	111
I.3.3. Réponses pour le plan d'optimisation	111
I.3.4. Géométrie du domaine expérimental	112
I.3.5. Modèle empirique	113
I.3.6. Construction du plan d'expériences	113
I.3.7. Réalisation des expériences	117
I.3.8. Analyse globale des résultats d'essais	120
I.3.9. Analyse mathématique des résultats d'essais	120
I.4. Analyse statistique des résultats.....	125
I.4.2. Analyse graphique des résultats	127
I.4.3. Conclusion du plan d'optimisation	130
I.5. Analyses de données.....	132
I.5.1. Analyse 1 : propriétés des films	132
I.5.2. Analyse 2 : propriétés des films et plasma	140
I.6. Conclusion.....	146

Chapitre II : Optimisation des couches minces de titanate de baryum et de strontium obtenues par ablation laser.....147

II.1. Formalisation du problème.....	147
II.1.1. Objectif	147
II.1.2. Réponses	149
II.1.3. Facteurs	149
II.2. Analyse de données existantes.....	150
II.2.1. Tableau de données initiales	150
II.2.2. Pré-traitement des données	153
II.2.3. Analyse en composantes principales	155
II.2.4. Classification ascendante hiérarchique	160
II.2.5. Conclusion sur l'analyse de données	160
II.3. Optimisation.....	161

II.3.1. Objectifs de l'optimisation	161
II.3.2. Facteurs –niveaux-domaine expérimental	161
II.3.3. Modèle empirique	162
II.3.4. Construction du plan d'expériences	162
II.4. Analyse Factorielle Discriminante.....	164
II.5. Conclusion.....	166

Chapitre III : Application à la caractérisation hyperfréquence de couches minces

III.1. Formalisation du problème.....	167
III.1.1. Objectif	167
III.1.2. Réponses	169
III.1.3. Facteurs	170
III.2. Etape 1 : paramètres EM du circuit à vide.....	170
III.2.1. Description du plan d'expériences de l'étude 1	170
III.2.2. Analyses et résultats	171
III.3. Etape 2 : paramètres EM du substrat.....	178
III.3.1. Description du plan d'expériences de l'étude 2	178
III.3.2. Analyses et résultats	179
III.4. Etape 3 : paramètres EM de la cm.....	179
III.4.1. Description du plan d'expériences de l'étude 3	179
III.4.2. Analyses et résultats	180
III.5. Conclusion.....	181

CONCLUSION GENERALE.....	183
--------------------------	-----

ANNEXE

Annexe 1.....	187
Annexe 2.....	196
Annexe 3.....	200
Annexe 4.....	203

REPERTOIRE.....	206
-----------------	-----

Introduction générale

Depuis une vingtaine d'années, les statistiques ont évolué dans différentes directions, parmi lesquelles l'analyse de données et les plans d'expériences.

L'analyse de données permet d'analyser de grands ensembles de données non structurés, notamment quand les paramètres ne peuvent pas être suffisamment maîtrisés. Ces méthodes d'analyses permettent d'interpréter des essais déjà réalisés et de décrire les influences des paramètres mis en jeu, de manière qualitative. Par exemple, tel paramètre sera fortement ou pas influent sur une réponse d'un système. Sous le terme analyses de données, sont regroupées l'ensemble des méthodes permettant de collecter, d'organiser, de résumer, de présenter et d'étudier des données de façon à en tirer le maximum d'informations.

La méthodologie des plans d'expériences permet une recherche expérimentale planifiée appelée « plans d'expériences ». L'expérimentation ne peut pas être quelconque : elle doit fournir l'information désirée. Cette démarche expérimentale va aider l'expérimentateur à structurer sa recherche de manière différente, à confronter et à valider ses propres hypothèses, à mieux comprendre les phénomènes étudiés et à solutionner les problèmes. Le succès de cette méthodologie est en partie lié aux besoins de compétitivité des entreprises mais aussi, à une envie de changer la manière de faire des expérimentations.

A cela, il faut ajouter les techniques d'aide à la formulation d'un problème qui mettent en évidence l'importance prépondérante de l'étape de « Réflexion » qui devrait précéder toute expérimentation.

Ce manuscrit se subdivise en trois grandes parties.

- La première partie (**Partie A**) traite de la théorie sur les plans d'expériences. Dans cette partie, après avoir montré l'intérêt des plans d'expériences par rapport à la méthode classique des essais et des erreurs, nous décrivons les outils (brainstorming, grille de choix des facteurs, diagramme d'Ishikawa) nécessaires à la formalisation d'un problème. Puis, nous présentons les deux types de stratégies expérimentales que nous avons choisies d'utiliser, à savoir, les plans de criblage et les plans à surfaces de réponses.
- La deuxième partie (**Partie B**) porte sur la théorie des méthodes d'analyses de données. Les outils de prétraitement des données tels que les méthodes d'analyses unidimensionnelles (Box plot) et bidimensionnelles (matrice des corrélations, nuage de points) sont tout d'abord décrites. Après avoir mis en évidence l'importance du prétraitement des données, trois méthodes d'analyses multidimensionnelles sont abordées. L'Analyse en Composante Principale qui décrit les corrélations entre tous les paramètres quantitatifs (entrées et sorties

d'un système), la Classification Ascendante Hiérarchique qui répartit l'ensemble des expériences en groupes homogènes (familles d'expériences) et l'Analyse Factorielle Discriminante qui permet de corréler un paramètre qualitatif à un ensemble de paramètres quantitatifs.

- La troisième partie est consacrée au développement d'exemples associés aux méthodes d'analyses de données et aux plans d'expériences.

Le premier exemple porte sur l'optimisation de couches minces d'alumine réalisées par P.E.C.V.D. A travers cet exemple, nous expliquons la construction et la démarche des plans de criblage et des plans à surfaces de réponses ainsi, que l'application d'une méthode d'analyses multidimensionnelles de données : l'Analyse en Composantes Principales.

L'optimisation de couches minces de titanate de baryum et de strontium obtenues par ablation laser, constitue le deuxième exemple. Dans ce cas, nous avons détaillé l'étape de prétraitement des données (analyses unidimensionnelles et bidimensionnelles) et développé trois méthodes d'analyses multidimensionnelles : l'Analyse en Composantes Principales, la Classification Ascendante Hiérarchique et l'Analyse Factorielle Discriminante.

Dans le troisième exemple, nous avons utilisé des plans à surfaces de réponses et développé plus particulièrement une méthode d'optimisation multi-critères (désirabilité) appliquée à la caractérisation hyperfréquence de couches minces.

Enfin, une étude en cours sur l'optimisation des propriétés du niobate de lithium massif sous irradiation par faisceau d'électrons pilotés, est abordé dans l'**annexe 1** de ce manuscrit.

Partie A :

Rappels théoriques sur les plans d'expériences

Introduction

Grâce au déploiement de la qualité dans les entreprises et à un accès facilité aux moyens de calcul, les plans d'expériences prennent une place plus importante. Les applications des plans d'expériences deviennent de plus en plus nombreuses. La méthodologie des plans d'expériences permet une recherche expérimentale planifiée appelée « plans d'expériences ». Elle va aider l'expérimentateur à structurer sa démarche expérimentale de manière différente (voir **Chapitre I**), à confronter et à valider ses propres hypothèses, à mieux comprendre les phénomènes étudiés et à résoudre les problèmes.

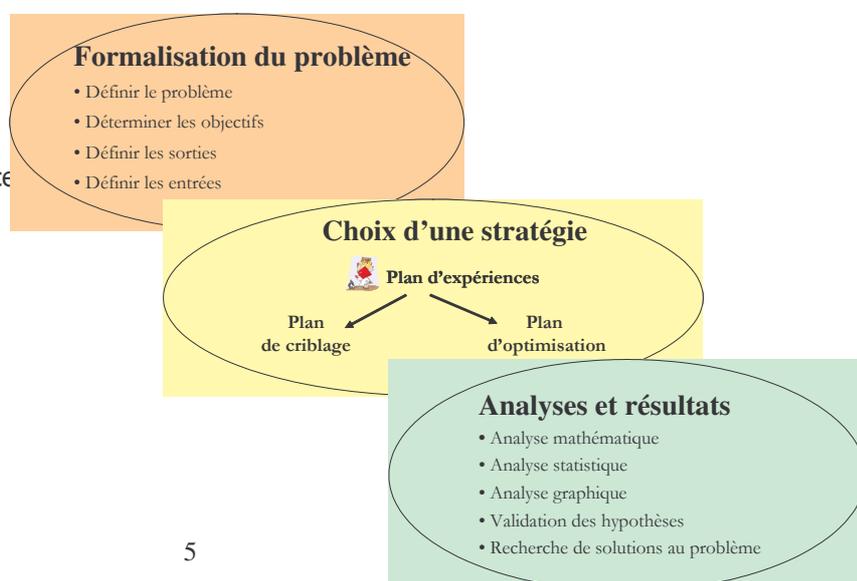
Le succès des plans d'expériences est en parti lié aux besoins de compétitivité des entreprises mais aussi, à une envie de changer la manière de faire les expérimentations. Actuellement, l'image du chercheur solitaire tenant compte de son « flair » pour diriger sa recherche a vieilli, les chercheurs et les ingénieurs reconnaissent donc la nécessité d'une démarche rationnelle [1].

Dans cette partie, après avoir montré l'intérêt des plans d'expériences (**Chapitre I**), nous aborderons la méthodologie à mettre en œuvre pour les utiliser. Cette méthodologie s'articule selon trois étapes (voir **Figure A.1**) :

- la **formalisation du problème** expérimental c'est-à-dire les étapes préliminaires qui posent la problématique (**Chapitre II**),
- si l'analyse du problème conduit à choisir une **stratégie** de construction d'un plan d'expériences alors, l'expérimentateur a accès à deux familles de plans : les plans pour l'étude des facteurs (plans de criblage, ...) et les plans pour la modélisation des surfaces de réponses (plans d'optimisation) (**Chapitre III**) qui dépendront d'un modèle empirique postulé à priori,
- les **analyses** (statistiques, mathématiques et graphiques) et les **résultats** (interprétations, validations, conclusions) détaillés dans le **Chapitre IV**.

Figure A.1

Mise en œuvre d'un plan d'expérience



Chapitre I : Origine et état de l'art

I.1. POURQUOI FAIRE DES PLANS D'EXPERIENCES ?

Les progrès des sciences et de la technologie sont étroitement liés aux réponses que l'homme a pu apporter aux interrogations que lui pose régulièrement la nature. Ces réponses résultent le plus souvent d'une analyse des observations expérimentales obtenues par une approche méthodologique rigoureuse. Nous pouvons donc dire que, sans expérimentation physique ou numérique, il n'y a pas de solution. L'expérimentation est donc un des moyens privilégiés pour acquérir ou améliorer les connaissances mais, elle doit être optimisée car l'objectif est d'obtenir des informations les plus fiables possibles en un minimum d'essais. Dans ce paragraphe, nous allons nous intéresser aux deux méthodes d'expérimentations couramment utilisées dans les laboratoires. C'est au regard des limites de ces méthodes que les plans d'expériences ont été créés.

I.1.1. LA RECHERCHE EXPLORATOIRE⁽¹⁾

Au tout début d'une nouvelle étude, le chercheur ignore souvent presque tout du phénomène qu'il va étudier. Il ignore même parfois s'il va réussir à le produire. Il conduit alors des expériences préliminaires destinées à cerner un ou plusieurs domaines expérimentaux, à s'assurer de la maîtrise du phénomène, à choisir les grandeurs qui lui serviront de réponses ultérieurement et à en contrôler la répétabilité.

I.1.2. METHODE TRADITIONNELLE DES ESSAIS - ERREURS

I.1.2.a. Démarche

Traditionnellement, les essais sont effectués de manière séquentielle en faisant varier les facteurs d'entrée du système étudié l'un après l'autre sans planification préalable de l'ensemble des essais à réaliser. Il en résulte que la décision d'arrêter ou de continuer les essais, repose essentiellement sur des résultats obtenus pas à pas. L'expérimentateur tente un premier essai, en tire des conclusions qui vont induire un deuxième essai, et ainsi de suite, en faisant souvent varier un seul facteur à la fois entre deux essais consécutifs. Cette méthode traditionnelle, appelée essais - erreurs, est schématisée sur la **Figure A.2**. L'interprétation des résultats se fait donc pas à pas, en comparant le résultat de l'essai effectué avec un résultat précédent, afin d'en déduire l'effet sur la réponse du paramètre dont le niveau vient d'être modifié.

Cependant, compte tenu de la très grande complexité des phénomènes mis en jeu dans certains procédés, le développement par une méthode essais - erreurs est parfois inenvisageable.

⁽¹⁾Note de stage, *Méthodologie de la recherche expérimentale MRE, S.A.R.L. LPRAI, MARSEILLE.*

Il est alors indispensable d'avoir recours aux outils de connaissance scientifique structurés sous forme de modélisation multidimensionnelle. Dans ce cas, la méthode s'appuie sur l'expérimentation mais, on ne laisse pas le hasard choisir les expériences. Nous verrons par la suite, que ces outils permettent de « fabriquer » de l'expérience et d'en tirer de manière claire et simple de l'information.

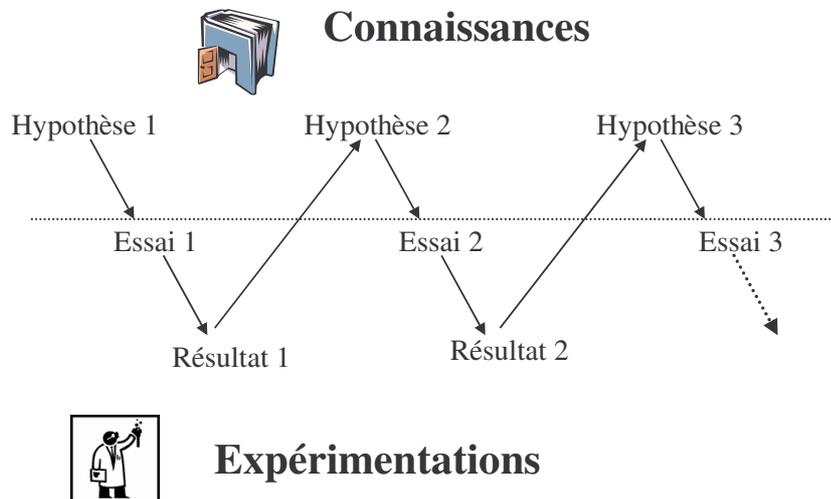


Figure A.2 Méthode des essais et des erreurs

I.1.2.b. Limites des méthodes traditionnelles d'essais-erreurs

Les essais classiques ne permettent pas toujours d'optimiser le produit ou le processus. En effet, le principe même des essais conduit le plus souvent à emprunter un chemin labyrinthique et à s'arrêter dans le meilleur cas à la première solution jugée satisfaisante. Par suite, la solution retenue n'est pas nécessairement optimale dans la mesure où d'autres solutions non explorées, auraient peut-être pu permettre d'atteindre de meilleures performances ou, pour des résultats identiques, d'obtenir une solution plus économique. Cette méthode ne met pas forcément en lumière les interactions entre les différents paramètres. Certains effets ainsi calculés en comparant deux à deux les résultats d'essais, ne sont valables que pour des valeurs particulières des paramètres et ne peuvent souvent se transposer à d'autres conditions d'essais. De plus, les expérimentations selon les méthodes classiques ne permettent pas de tenir compte facilement de la variabilité des grandeurs observées, occasionnées par une multitude de sources de fluctuation.

En définitive, ces méthodes :

- conduisent à des conclusions nécessitant des remises en cause fréquentes,
- impliquent une durée de campagne expérimentale parfois longue,
- donnent une connaissance très relative des influences des facteurs,

- ne permettent pas de prévoir le comportement du système étudié pour des valeurs non testées des paramètres,
- manquent un peu de formalisme, de rigueur et de méthode pour capitaliser et valoriser l'expérience technique.

I.2. METHODOLOGIE DES PLANS D'EXPERIENCES

I.2.1. QU'EST CE QU'UN PLAN D'EXPERIENCES ?

Les plans d'expériences constituent essentiellement une stratégie de planification d'expériences afin d'obtenir des conclusions solides et adéquates de manière efficace et économique. La méthodologie des plans d'expériences se base sur le fait qu'une expérience convenablement organisée, conduira fréquemment à une analyse et à une interprétation statistique relativement simple des résultats (Norme ISO 3534-3).

I.2.2. HISTORIQUE ET ETAT DE L'ART

La méthode des plans d'expériences est à la fois nouvelle et ancienne [2]. Nouvelle pour l'ingénierie mécanique et électronique, qui traditionnellement, est portée par des sciences exactes. On prend conscience aujourd'hui que loin de nier la connaissance technique, la méthode des plans d'expériences la valorise. Initialement développés et mis en pratique par le mathématicien britannique **Sir Ronald Fisher** [3], au début des années 30, les plans d'expériences ont servi à quantifier l'effet des facteurs contrôlés par l'expérimentateur dans un contexte soumis à de nombreuses sources d'hétérogénéité. Après 1945, les plans d'expériences suscitent de nombreuses publications et recherches dans le monde anglo-saxon. Des statisticiens comme **Yates**, **Cochran**, **Plackett** et **Burmann**, enrichissent et divulguent la méthode [4][5][6]. **Box** et **Hunter** [7], s'appuyant sur les travaux de **Yates** [5], développent des méthodes particulières de construction de plans d'expériences fractionnaires à deux niveaux (criblage). Ils s'intéressent à l'étude des facteurs quantitatifs et introduisent, dans les années 60, les modèles de surface de réponses associés à des plans comme des plans composites centrés.

Cependant, à l'époque, seuls des statisticiens spécialistes pouvaient mettre en œuvre ces méthodes. A partir des années 50, en s'attaquant à l'amélioration de la qualité, le Japon imprime un nouveau souffle. **Taguchi** et **Masuyama** élaborent des tables permettant de construire des plans d'expériences adaptés à la majorité des problèmes industriels [8][9]. Le mérite de **Taguchi** est d'avoir contribué à une méthodologie facile à appliquer. Initialement difficiles d'accès, ces méthodes se laissent adapter, vulgariser pour conquérir un cercle d'utilisateurs de plus en plus large.

Elles trouvent donc une nouvelle jeunesse en encourageant le travail de groupe, la mise en commun et la synergie des savoirs et du savoir faire, et la réflexion préalable avant l'action.

Les applications des plans d'expériences ont tendance actuellement à se diversifier ; on voit donc l'émergence de nouveaux plans. Leur diffusion à grande échelle dans les entreprises, leur enseignement et leurs applications dans les universités en sont encore à leurs débuts. L'effort amené par **Taguchi** se poursuit et des logiciels de plus en plus conviviaux destinés à la construction et à l'analyse de plans d'expériences se développent.

Chapitre II : Formalisation du problème

II.1. INTRODUCTION

L'étape préliminaire de formalisation du problème nécessite de regrouper les acteurs d'une étude autour d'un **Brainstorming**. Le **Brainstorming** ou « Remue-méninges » est une technique de créativité de groupe permettant de produire le plus d'idées possibles, dans un minimum de temps, sur un thème donné. Cette technique est utilisée dans la plupart des étapes de la résolution de problèmes, notamment pour recenser les questions suscitées lors du lancement d'un projet ou pendant son déroulement, pour en déterminer les causes possibles et trouver toutes les solutions ou éléments de solution possibles.

Le brainstorming doit être organisé par un animateur qui doit annoncer le but recherché, s'assurer de la représentativité des personnes convoquées et animer le groupe en favorisant la production d'idées.

Le déroulement d'un brainstorming fait appel aux trois points suivants :

- la présentation et la formulation du problème sous forme d'une question,
- la production d'idées : les membres du groupe énoncent toutes les idées sur le sujet leur passant par la tête jusqu'à épuisement des idées,
- l'exploitation des idées : on reformule les idées peu claires en éliminant celles qui sont hors sujet afin de regrouper les idées similaires puis de hiérarchiser les idées finales. L'objectif est de mettre en évidence les priorités recherchées qu'il s'agisse des causes ou des solutions au problème posé.

La formalisation du problème permet de rassembler l'expérience accumulée de manière diffuse et de la mettre dans un « moule » afin de la compléter et de la préciser grâce à l'expérimentation. Plus cette étape de formalisation sera établie avec soin, plus les plans d'expériences seront fructueux.

- En s'appuyant sur cette méthode, nous allons : définir le problème, déterminer les objectifs de l'étude et faire ressortir les paramètres d'entrée ainsi que les variables de sortie du système étudié.

II.2. DEFINIR LE PROBLEME

Définir le problème, c'est le poser clairement en précisant de quoi il s'agit, et en quoi le sujet proposé est un problème. Pour cela, il est recommandé de faire appel à la méthode du QQCOQP qui consiste à répondre aux questions suivantes [1].

Q : En **Quoi** consiste le problème ?

Q : **Qui** est gêné par le problème ? Qui est le demandeur de l'étude ?

C : **Combien** de cas ? Combien ça coûte ?

O : **Où** cela se passe-t-il ?

Q : **Quand** cela arrive-t-il ?

P : **Pourquoi** est-ce un problème ?

II.3. DETERMINER LES OBJECTIFS

Pour formuler le plus clairement possible les objectifs visés, il est conseillé d'écrire une phrase courte et précise définissant clairement le but visé. De manière générale, les deux objectifs principaux visés sont :

- « Débroussailler », c'est-à-dire faire le tri parmi l'ensemble des variables d'entrées susceptibles d'influencer les variables de sortie et les hiérarchiser. Ce sera le cas s'il existe un grand nombre de variables d'entrées. Il peut s'agir d'une première étape conduisant à mener d'autres analyses pour avoir plus de précision sur le réglage de ces variables influentes.
- « Prévoir », c'est-à-dire estimer avec suffisamment de précision l'effet de l'ensemble des variables d'entrées influentes, de façon à pouvoir modéliser de manière assez fiable, la valeur de la variable de sortie que l'on obtiendrait pour l'ensemble des combinaisons des variables d'entrées, y compris celles non essayées. Ceci est notamment nécessaire si l'on souhaite optimiser le système étudié.
- Il est également indispensable de recenser les contraintes de l'étude comme le nombre d'essais maximum qu'il sera possible de réaliser et les ressources disponibles.

II.4. DEFINIR LES SORTIES (REponses)

La **réponse Y** (variable de sortie) correspond à un paramètre de sortie du système étudié. Une réponse doit être représentative, quantifiable et la moins dispersée possible pour des variables d'entrées maîtrisées et constantes. Pour appliquer la méthodologie des plans d'expériences, il est conseillé d'avoir une réponse exprimée sous forme quantitative. En effet, les méthodes d'analyses des résultats d'essais telles que l'analyse de variance ou l'analyse de régression au sens des moindres carrés, s'appuient sur des données exclusivement quantitatives. Par ailleurs, les réponses qualitatives comme, par exemple, la morphologie d'un dépôt (lisse, poreux...etc.) qui n'est pas un critère quantitatif, pourront s'interpréter à partir d'un codage spécifique des modalités des réponses.

II.5. DEFINIR LES ENTREES (FACTEURS)

Un paramètre d'entrée d'un système (variable d'entrée) est appelé **facteur X**. Ce facteur est une cause possible de variation de la réponse Y. Le choix d'un facteur se fait souvent en plusieurs phases : une phase de recensement faisant appel à la créativité du groupe, une phase de classement et une phase de sélection où les modalités des facteurs devant être étudiés de façon prioritaire seront définies.

II.5.1. PHASE DE CREATIVITE

Le groupe de travail établit une liste de facteurs potentiels. Le diagramme causes - effets ou diagramme d'Ishikawa (voir **Figure A.3**) est une aide particulièrement précieuse pour structurer la présentation des facteurs par grandes familles. Ce diagramme se construit avec tous les acteurs d'une étude.

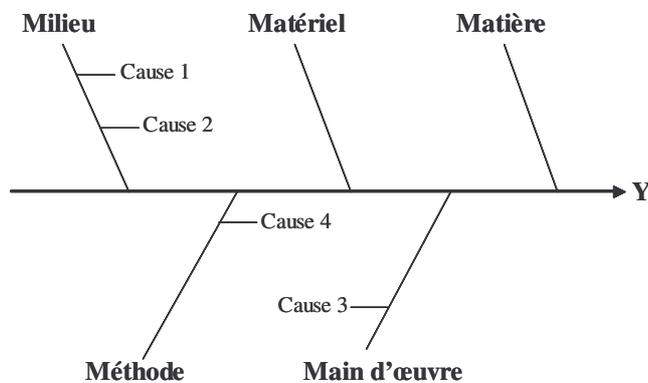


Figure A.3 Diagramme d'Ishikawa

Chacune des causes sera délimitée par un domaine expérimental que nous formaliserons avec l'expérimentateur.

II.5.2. PHASE DE CLASSEMENT [10]

La phase de classement consiste à déterminer quels sont les facteurs puis, grâce à une grille de choix de les classer en fonction de leur importance compte tenu des priorités de l'étude.

II.5.2.a. Catégories de facteurs

Un facteur est une variable de prédiction qui varie en vue de l'évaluation de son effet sur la variable de réponse. Un facteur peut aussi fournir une cause assignable aux résultats d'une expérience. Un facteur peut-être associé à la création de blocs (ISO 3534-3).

- variables de prédiction : variable susceptible de contribuer à l'explication d'un résultat d'une expérience. Le degré auquel une variable de prédiction donnée peut être maîtrisée, régit son rôle potentiel dans un plan d'expériences.

Les variables de prédiction sont susceptibles d'être maîtrisées (fixes), modifiables (maîtrisées uniquement durant une courte période ou à un coût considérable) ou non maîtrisables (aléatoires) (ISO 3534-3).

- **Bloc** : groupement d'unités expérimentales plus homogènes que l'ensemble complet des unités expérimentales. Le terme de « bloc » provient des expériences agronomiques dans lesquelles un champ est subdivisé en sections présentant des conditions communes telles que exposition au vent, proximité d'eau souterraine ou épaisseur de la couche arable. Dans d'autres conditions, les blocs sont constitués par des lots de matières premières, des opérateurs, le nombre d'unités étudiées dans une même journée, etc. En général, le fait de reconnaître l'existence des blocs peut affecter la manière dont les traitements sont affectés aux unités expérimentales (ISO 3534-3).

II.5.2.b. Grille de choix des facteurs ⁽¹⁾

Dans le cas où l'expérimentateur ne peut effectuer qu'un nombre restreint d'expériences, on construit une grille de choix des facteurs principaux afin de déterminer pour chaque facteur, un **indice de priorité**. Pour chacun des facteurs, on définit une note comprise entre 1 et 10 pour quantifier l'efficacité supposée d'un facteur sur une réponse, la faisabilité à court terme des expériences et le besoin d'information nécessaire. L'indice de priorité est obtenu par le produit des notes (voir **Tableau A.1**), il permet alors de justifier le choix des facteurs à étudier. Dans notre exemple, le facteur A est supposé efficace (8), peu d'information est connue (10) sur ce facteur, il semble également très difficile à étudier à court terme (1) ce qui donne un indice de priorité faible (80). Par ailleurs, nous avons beaucoup d'information sur le facteur C (1), celui-ci est moyennement facile à étudier à court terme (5), on obtient donc un indice de priorité très faible (50). Ce sera donc le facteur B qui sera retenu pour une étude à court terme parmi ces trois facteurs avec un indice de priorité égale à 441.

Tableau A.1 Exemple de grille de choix des facteurs

Facteurs	Efficacité supposée	Faisabilité à court terme	Besoin d'information	Indice de priorité
A	8	1	10	80
B	7	9	7	441
C	10	5	1	50

⁽¹⁾Note de cours, François Louvet, *Plans d'expériences - Approche méthodologique*, 2003

II.5.2.c. Interactions pressenties

Nous allons dans un premier temps définir quelques termes :

Interaction : effet pour lequel l'influence apparente d'un facteur sur une variable de réponse dépend d'un ou de plusieurs facteurs. L'interaction indique une incohérence de l'effet principal d'un facteur sur la réponse selon le niveau d'un autre facteur (ISO 3534-3).

Effet moyen d'un facteur : variation de la *réponse* observée ou modélisée lorsque le *facteur* change de *modalité* (définition de l'association « Expérimentative »).

Il est également possible de faire un tableau intuitif des interactions pressenties par les expérimentateurs comme l'indique le **Tableau A.2**. Dans cet exemple, l'expérimentateur pressent une interaction entre les facteurs A et B.

	A	B	C
A			
B	↕		
C			

Tableau A.2 Interactions pressenties

A ce stade, nous avons défini la liste des facteurs à faire varier et celle des facteurs à fixer selon les limites économiques. Fixer un facteur consiste à lui imposer la même valeur dans tous les essais liés à l'incertitude de réglage de celui-ci. On ne pourra pas l'analyser mais on évitera que ces variations ne perturbent les résultats d'essais, on supposera que les variations des incertitudes sur ces facteurs ont peu de conséquence sur la variation de la réponse.

II.5.3. DEFINITION DES FACTEURS

Après avoir déterminé le nombre **k** de facteurs principaux d'un système, il faut maintenant décider, pour chaque facteur, quelles valeurs ils sont susceptibles de prendre durant l'expérimentation. Deux cas sont à considérer selon la nature de ces facteurs : facteurs qualitatifs ou quantitatifs.

II.5.3.a. Facteurs qualitatifs

Un facteur qualitatif prend ou on lui attribue des **modalités (m_i)** qui sont soit non numériques (couleurs, descriptions, ...), soit numériques et discrètes (appréciations, dosage d'une poudre,...). Le nombre et le choix des modalités d'un facteur qualitatif sont laissés à l'expérimentateur. Ces facteurs sont destinés à l'étude des effets qui s'effectue lorsqu'il y a nécessité de faire un choix (à cause des limites économiques) entre les différents facteurs. Il faudra donc justifier le nombre et le choix des modalités en fonction des besoins d'information recherchée en terme d'effet des facteurs.

II.5.3.b. Facteurs quantitatifs

Un facteur quantitatif prend des valeurs continues ou discrètes appelées **niveaux**. Le choix du nombre de niveaux et la valeur prise U_i par chaque facteur à chacun des niveaux est défini par le plan d'expériences choisi et l'étendue de ces facteurs. L'étendue de variation de ces facteurs est matérialisée à partir de deux limites U_{\min} et U_{\max} à l'intérieur desquelles on définira les niveaux U_i imposés par le plan d'expériences. Il est important de s'informer sur la difficulté de réglage de ces facteurs dans les plages données et intéressant de savoir s'il est possible de faire des expériences en dehors du domaine expérimental (défini dans le paragraphe suivant).

Pour un facteur donné, qualitatif ou quantitatif, le **nombre de modalités** ou de **niveaux** sera défini par s_i .

II.5.4. DOMAINE EXPERIMENTAL

II.5.4.a. Domaine expérimental continu

Un domaine expérimental continu est délimité par des facteurs quantitatifs. La **Figure A.4** illustre le domaine expérimental de deux facteurs quantitatifs 1 et 2. Le facteur 1 varie de 15 à 50 et le facteur 2 varie de 10 à 40. Le nombre total de combinaisons que l'on peut réaliser est presque infini dans ce domaine.

Il est également possible, comme le montre également la figure, de restreindre le domaine expérimental en appliquant des contraintes. Ces contraintes peuvent être dues à des expériences impossibles à mettre en œuvre ou encore à des combinaisons de facteurs improbables.

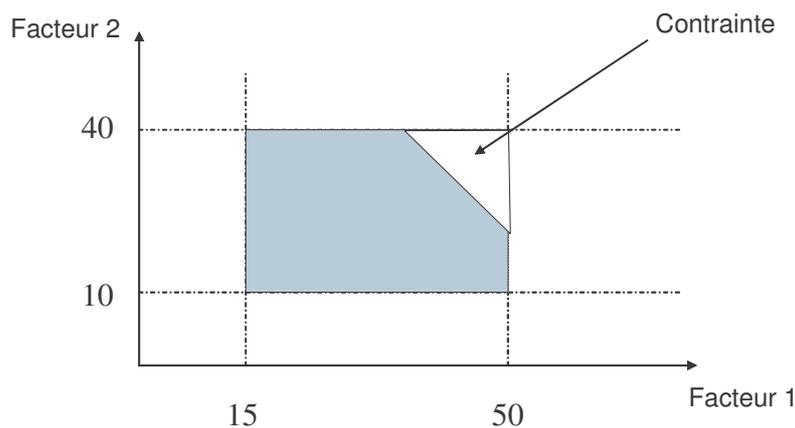


Figure A.4 Domaine expérimental continu

II.5.4.b. Domaine expérimental discret

Un domaine expérimental discret est délimité par des facteurs discrets ou qualitatifs. La **Figure A.5** illustre le domaine expérimental de deux facteurs qualitatifs A et B. Les facteurs A et B prennent chacun trois modalités A1, A2, A3 pour A et B1, B2, B3 pour B.

Dans ce cas, nous sommes donc en présence d'un nombre d'expériences réalisables limitées. Dans notre exemple, nous pouvons dire qu'il existe 9 expériences réalisables.

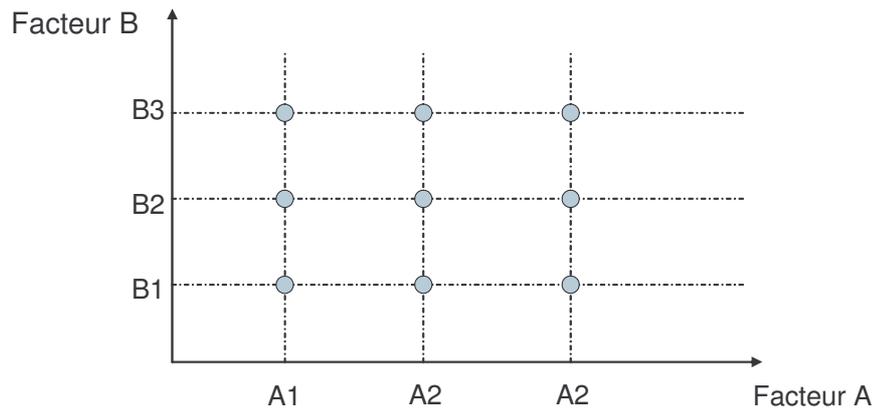


Figure A.5 Domaine expérimental discret

Après avoir formalisé le problème, il faut maintenant choisir la stratégie expérimentale à mettre en œuvre.

Chapitre III : Choix d'une stratégie

Nous allons expliquer dans ce chapitre, pourquoi les stratégies des plans d'expériences sont des stratégies expérimentales efficaces. Ensuite, nous ferons une introduction aux deux stratégies que nous allons étudier : les plans de criblage et les plans d'optimisation.

III.1. L'EFFICACITE DES PLANS D'EXPERIENCES

La méthodologie des plans d'expériences semble être une méthode d'expérimentation avantageuse. Voyons ses avantages par rapport aux différentes méthodes d'expérimentations que nous avons envisagées dans le paragraphe précédent. Pour cela, nous allons aborder cette étude au travers d'un exemple, celui de l'étude de l'effet de 7 facteurs ($X_1, X_2, X_3, X_4, X_5, X_6, X_7$) pouvant prendre chacun 2 valeurs (modalités), que nous noterons -1 et +1 par convention.

III.1.1. STRATEGIE TRADITIONNELLE : « UN SEUL FACTEUR A LA FOIS »

Elle consiste à faire varier successivement chacun des facteurs à chaque expérience, tous les autres étant maintenus constants. Cela peut conduire par exemple à la matrice d'expériences du **Tableau A.3** en considérant l'état de référence au numéro d'essai 1, c'est-à-dire lorsque tout les facteurs sont au niveau -1.

On voit donc que ce plan nécessite 8 expériences. Les résultats obtenus permettent d'estimer l'effet de chaque facteur de la manière suivante :

$$a = \text{effet du facteur } X_1 = Y_2 - Y_1$$

$$b = \text{effet du facteur } X_2 = Y_3 - Y_1 \dots \text{etc.}$$

Tableau A.3 Matrice d'expériences - Plan à un seul facteur à la fois

Numéro d'essai	X_1	X_2	X_3	X_4	X_5	X_6	X_7	Réponse
1	-1	-1	-1	-1	-1	-1	-1	Y_1
2	+1	-1	-1	-1	-1	-1	-1	Y_2
3	+1	+1	-1	-1	-1	-1	-1	Y_3
4	+1	+1	+1	-1	-1	-1	-1	Y_4
5	+1	+1	+1	+1	-1	-1	-1	Y_5
6	+1	+1	+1	+1	+1	-1	-1	Y_6
7	+1	+1	+1	+1	+1	+1	-1	Y_7
8	+1	+1	+1	+1	+1	+1	+1	Y_8

Avec cette méthode, l'effet de chacun des facteurs est calculé pour une combinaison particulière des autres facteurs et rien ne permet de dire si cet effet est également valable dans les autres cas (problème d'interaction entre facteurs). On parle alors d'effet conditionnel.

Il existe en fait autant d'effets conditionnels possibles qu'il y a de combinaisons envisageables pour les autres facteurs et le choix de la combinaison testée est tout à fait arbitraire : l'effet mesuré est donc peu représentatif de l'effet réel recherché, ce qui peut conduire à des aberrations dans l'exploitation des résultats des expériences.

III.1.2. STRATEGIE EXHAUSTIVE : LE PLAN FACTORIEL COMPLET

Le plan factoriel complet combine de manière exhaustive toutes les combinaisons possibles des variables avec un nombre de facteurs finis (facteurs qualitatifs). Si on considère un nombre discret de variation par facteurs, il permet d'envisager toutes les combinaisons possibles des modalités des facteurs N_{total} .

Nous allons considérer k facteurs X_1, X_2, \dots, X_k pouvant prendre chacun s_1, s_2, \dots, s_k modalités différentes. Par définition, une matrice d'expériences factorielle complète est constituée de toutes les combinaisons possibles des modalités prises par chaque facteur.

Dans le cas d'une **matrice d'expériences asymétrique** c'est-à-dire une matrice où les facteurs possèdent des modalités différentes, le nombre N_{total} de combinaisons distinctes est égal au produit :

$s_1^{k_1} s_2^{k_2} \dots s_i^{k_i}$ (avec k_1 : nombre de facteurs contenant s_1 modalités, k_2 : nombre de facteurs contenant s_2 modalités, k_i : nombre de facteurs contenant s_i modalités ...).

Prenons un exemple où $k = 5$ facteurs qualitatifs (X_1, X_2, X_3, X_4, X_5).

X_1 est un facteur à 2 modalités ($s_1 = 2$)

X_2 est un facteur à 2 modalités ($s_2 = 2$)

X_3 est un facteur à 2 modalités ($s_3 = 2$)

X_4 est un facteur à 3 modalités ($s_4 = 3$)

X_5 est un facteur à 3 modalités ($s_5 = 3$)



$k_1 = 3$ (3 facteurs à 2 modalités)

$k_2 = 2$ (2 facteurs à 3 modalités)

$N_{\text{total}} = 2^3 \times 3^2 = 72$ combinaisons

Pour une **matrice d'expériences symétrique**, c'est-à-dire dans laquelle tous les facteurs (k) ont le même nombre de modalités (s), le nombre total (N_{total}) de combinaisons distinctes est égal à s^k .

Pour le plan précédent à 7 facteurs à 2 modalités, le plan factoriel complet comprendra $2^7 = 128$ expériences à réaliser (voir **Tableau A.4**). L'intérêt du plan factoriel complet est, bien sûr, de permettre d'estimer avec la même incertitude tous les effets des différents facteurs, et aussi de trouver directement la meilleure combinaison puisque toutes les combinaisons avec les modalités choisies sont essayées.

Tableau A.4 Plan factoriel complet, k = 7, s = 2

Numéro d'essai	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	Réponse
1	-1	-1	-1	-1	-1	-1	-1	Y ₁
2	-1	-1	-1	-1	-1	-1	+1	Y ₂
3	-1	-1	-1	-1	-1	+1	-1	Y ₃
4	-1	-1	-1	-1	-1	+1	+1	Y ₄
5	-1	-1	-1	-1	+1	-1	-1	Y ₅
6	-1	-1	-1	-1	+1	-1	+1	Y ₆
7	-1	-1	-1	-1	+1	+1	-1	Y ₇
8	-1	-1	-1	-1	+1	+1	+1	Y ₈
9	-1	-1	-1	+1	-1	-1	-1	Y ₉
10	-1	-1	-1	+1	-1	-1	+1	Y ₁₀
11	-1	-1	-1	+1	-1	+1	-1	Y ₁₁
12	-1	-1	-1	+1	-1	+1	+1	Y ₁₂
13	-1	-1	-1	+1	+1	-1	-1	Y ₁₃
14	-1	-1	-1	+1	+1	-1	+1	Y ₁₄
15	-1	-1	-1	+1	+1	+1	-1	Y ₁₅
16	-1	-1	-1	+1	+1	+1	+1	Y ₁₆
17	-1	-1	+1	-1	-1	-1	-1	Y ₁₇
...
127	+1	+1	+1	+1	+1	+1	-1	Y ₁₂₇
128	+1	+1	+1	+1	+1	+1	+1	Y ₁₂₈

Mais, le principal défaut réside bien sûr dans le nombre très élevé d'expériences que ce plan nécessite, défaut que l'on essaiera de contourner grâce à des plans d'expériences plus adaptés. Ce plan ne peut donc être envisagé que pour des études comportant un petit nombre de facteurs avec peu de niveaux.

III.1.3. LA STRATEGIE DU PLAN FRACTIONNAIRE ORTHOGONAL

Le plan fractionnaire orthogonal est une fraction judicieusement choisie du plan factoriel complet qui permet de tester un grand nombre de facteurs avec un minimum d'expériences et d'en déduire l'effet moyen de chaque facteur avec une assez bonne précision. L'analyse des résultats du plan comportant alors 8 expériences seulement (**Tableau A.5**) pourra estimer l'effet moyen des 7 facteurs à 2 modalités précédemment cités.

L'effet du facteur X₁ (**a**) est estimé en faisant la différence entre les moyennes de la réponse lorsque le facteur prend la valeur 1 et la valeur -1, ce qui se traduit par :

$$a = \frac{Y_5 + Y_6 + Y_7 + Y_8}{4} - \frac{Y_1 + Y_2 + Y_3 + Y_4}{4}$$

Cet effet est estimé par rapport à l'état de référence -1. Il en est de même pour les autres facteurs.

Tableau A.5 Plan fractionnaire orthogonal

Numéro d'essai	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	Réponse
1	-1	-1	-1	-1	-1	-1	-1	Y ₁
2	-1	-1	-1	+1	+1	+1	+1	Y ₂
3	-1	+1	+1	-1	-1	+1	+1	Y ₃
4	-1	+1	+1	+1	+1	-1	-1	Y ₄
5	+1	-1	+1	-1	+1	-1	+1	Y ₅
6	+1	-1	+1	+1	-1	+1	-1	Y ₆
7	+1	+1	-1	-1	+1	+1	-1	Y ₇
8	+1	+1	-1	+1	-1	-1	+1	Y ₈

Ce plan comporte 8 expériences, comme le plan « un facteur à la fois ». Quelles sont donc leurs différences ?

- Cette fraction du plan complet n'a pas été choisie au hasard, elle possède une propriété particulière : l'**orthogonalité**. Pour expliquer ce terme, observons les colonnes X₁ et X₂. nous remarquons que, chaque niveau du facteur X₁ est associé le même nombre de fois à chaque niveau du facteur X₂ (le niveau -1 du facteur X₁ est associé 2 fois avec le niveau -1 et 2 fois avec le niveau 1 du facteur X₂). Cette propriété définit l'orthogonalité de ces 2 facteurs. L'avantage de l'orthogonalité est de pouvoir comparer des moyennes donc on a une équipartition des modalités des facteurs, on obtient des résultats comparables.
- Un plan dans lequel tous les couples de facteurs à deux modalités vérifient cette propriété sera dit orthogonal.
- Cette propriété a une conséquence fondamentale sur le calcul des effets des facteurs. Dans les expériences servant à évaluer l'effet du facteur X₁, le facteur X₂ varie, si bien que l'on pourrait craindre que ces variations perturbent cette évaluation. Or il n'en est rien car, pour chacune des valeurs de X₁, le facteur X₂ varie de la même façon, prenant 2 fois la valeur -1 et 2 fois la valeur 1. On voit donc ici la grande différence avec la stratégie « un seul facteur à la fois » : au lieu d'adopter la stratégie traditionnelle qui est de faire varier un facteur « toutes choses étant égales par ailleurs », on utilise la technique du « toutes choses variant également par ailleurs », qui permet de prendre en compte de nombreux facteurs tout en réduisant le nombre d'expériences.
- La formule permettant de calculer l'effet d'un facteur montre que toutes les expériences du plan concourent à ce calcul : l'effet calculé représente alors un effet moyen et l'on sait que les moyennes sont plus précises que les valeurs individuelles. Pour s'en convaincre, calculons l'incertitude sur le calcul de \mathbf{a} (σ_a).

On appelle y_i la valeur de la réponse expérimentale et \hat{y}_i la valeur de la réponse modélisée (calculée avec le modèle). On obtient :

$$y_i = \hat{y}_i + e_i \quad \text{avec } e_i : \text{erreur résiduelle ou écart or} \quad \mathbf{e}_i = \mathbf{d}_i + \mathbf{g}_i$$

Avec

\mathbf{d}_i : erreur d'ajustement ou systématique lié au modèle, appelé également *Lack of Fit*

\mathbf{g}_i : erreur aléatoire dû à σ^2 . (avec σ écart type de), appelé *Pure Error*

- Rappel :

L'estimation de l'incertitude permet de déterminer la qualité de l'information obtenue pour estimer la précision sur le calcul des effets des facteurs.

En effet, l'incertitude est distribuée suivant une loi de moyenne 0 et de variance σ^2 ce qui s'écrit : $\text{var}(\mathbf{e}_i) = \sigma^2$, les erreurs expérimentales doivent être indépendantes ce qui se traduit par :

$\text{cov}(e_2, e_1) = \sigma^2$. De plus, on sait que :

$\text{var}(y_1 - y_2) = \text{var}(y_2) + \text{var}(y_1) - 2\text{cov}(y_2, y_1) = \text{var}(y_2) + \text{var}(y_1)$ car la covariance entre y_1 et y_2 est nulle (les erreurs e_1 et e_2 étant indépendantes).

Avec le plan « un seul facteur à la fois », l'effet du facteur X_1 vaut : $a = Y_2 - Y_1$ ce qui donne

$\sigma_a = \sqrt{\sigma^2 + \sigma^2} = 1.4\sigma$, pour le plan fractionnaire, $a = \frac{Y_5 + Y_6 + Y_7 + Y_8}{4} - \frac{Y_1 + Y_2 + Y_3 + Y_4}{4}$, sachant

que $\text{var}(\mathbf{ax}) = \mathbf{a}^2\text{var}(\mathbf{x})$, on obtient $\sigma_a = \sqrt{\frac{4\sigma^2}{16} + \frac{4\sigma^2}{16}} = 0.7\sigma$ soit un écart type deux fois moindre.

De manière générale, on démontre qu'à nombre d'essais donnés, les plans orthogonaux conduisent à faciliter le calcul des effets. La qualité de l'information restituée est donc meilleure avec le plan orthogonal, ce qui est bien une des qualités recherchées d'un plan d'expériences. Ce résultat explique pourquoi les plans orthogonaux ont été aussi abondamment étudiés, notamment par **G.Taguchi [8]** et **Box et Hunter [7]**, qui ont d'ailleurs proposé des tables permettant de les construire. Mais aujourd'hui le critère d'orthogonalité pour les plans d'expériences n'est pas forcément un critère obligatoire, nous avons un autre moyen de voir si on obtient une information de qualité acceptable avec la détermination du facteur d'inflation que nous aborderons dans la suite de cette partie (**Chapitre IV**).

III.2. CONSTRUCTION DU PLAN D'EXPERIMENTATION

III.2.1. CHOIX DE LA STRATEGIE EXPERIMENTALE

Les plans d'expériences ont pour but d'apporter le maximum d'informations avec un minimum d'expériences. La méthodologie des plans d'expériences consiste à fabriquer un modèle approximatif qui apportera des éléments d'informations à l'expérimentateur encore appelé « modèle empirique ».

Pour cela, le choix du plan d'expériences à mettre en œuvre dépend du problème posé. Dans cette thèse, seuls deux types de plans seront abordés : les plans de criblage et les plans d'optimisation.

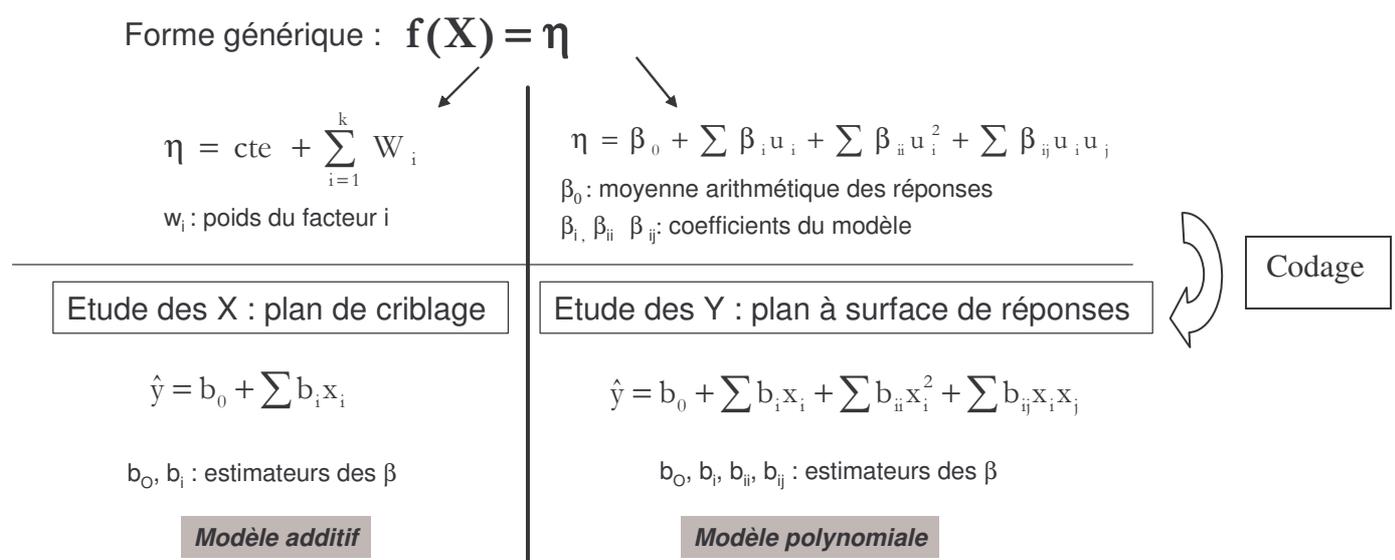
- Les **plans de criblage** permettent de déterminer le poids des facteurs sur les réponses d'un système à partir d'un modèle.

- Les plans **d'optimisation** ou **plans à surface de réponse** permettent à partir d'un modèle mathématique polynomial, de déterminer les valeurs des facteurs influents correspondant à une réponse particulière du système.

III.2.2. CHOIX DU MODELE

Pour l'ensemble des points constituant le domaine expérimental, nous voulons connaître la valeur de la réponse étudiée. Mais, pour des contraintes de coût, de temps, nous ne pouvons pas réaliser des expériences en tout point du domaine expérimental. C'est pourquoi, nous allons utiliser un **modèle empirique**, qui nous permettra d'avoir cette information en réalisant un minimum d'expériences [13].

Ce modèle est de la forme générale : **f(facteurs) = Réponses**



Ces fonctions polynomiales contiennent p coefficients ($b_0, b_1, b_2 \dots$ etc.) qu'il faudra estimer grâce à l'analyse des résultats d'essais. La nature et la qualité du modèle dépendent donc de ce que l'on souhaite établir comme relation entre les réponses et les facteurs.

III.2.3. CHOIX DE LA MATRICE D'EXPERIENCES

« Le choix du plan d'expériences dépend de la nature des questions à traiter, du degré de généralité recherché pour les conclusions et des ressources disponibles » (**Norme ISO 3534-3**)

III.2.3.a. Matrice d'expériences

Pour choisir un plan, il est important d'avoir clairement défini les questions pour trouver quels seront les moyens à mettre en œuvre pour y répondre par un plan d'expérimentation. Pour chacune des deux grandes orientations (criblage et optimisation), il existe un choix important de **matrices d'expériences**. Ces objets mathématiques codés permettent, après un recodage, d'obtenir un plan d'expérimentation à réaliser. Nous allons illustrer ci-dessous la notion de matrice d'expériences avec un exemple simple.

Prenons 2 facteurs U_1 et U_2 . U_1 est un facteur quantitatif qui peut prendre par exemple deux niveaux : $A_1 = 100$ et $A_2 = 50$; U_2 est un facteur qualitatif à deux modalités : $B_1 = \text{bleu}$ et $B_2 = \text{rouge}$. Supposons que la matrice d'expériences choisie soit celle reportée dans le **Tableau A.6** ci-dessous :

Tableau A.6 Matrice d'expériences

Numéro d'essai	U_1	U_2	Y
1	A_1	B_2	Y_1
2	A_2	B_2	Y_2
3	A_1	B_1	Y_3
4	A_2	B_1	Y_4

Ainsi on recode le Tableau A.6 pour obtenir le plan d'expérimentation à réaliser Tableau A.7.

Tableau A.7 Plan d'expérimentation

Numéro d'essai	X1	X2	Y
1	100	rouge	Y1
2	50	rouge	Y2
3	100	bleu	Y3
4	50	bleu	Y4

Dans le cas des **plans à surface de réponse**, les variables seront recodées avec la relation de codage ci-dessous :

$$x_i = \left[\frac{u_i - (u_{\max} + u_{\min})}{2} \right] / \left[\frac{u_{\max} - u_{\min}}{2} \right]$$

avec u_i : valeur de la variable U_i initiale.

x_i : valeur de la variable codée X_i .

Il existe un grand nombre de matrices d'expériences disponibles [12] mais, il est également possible de les construire pour pouvoir adapter le plan à un problème (voir **Partie C**).

III.2.4. LES CONTRAINTES PARTICULIERES A LA CONSTRUCTION D'UN PLAN

□ **LES ESSAIS INTERDITS :**

Il est important de préciser les incompatibilités physiques d'association de niveaux entre différents facteurs. La construction du plan devra en tenir compte pour ne pas créer des essais impossibles à réaliser.

□ **LES ESSAIS OBLIGATOIRES :**

Lorsqu'une pré-étude a été réalisée, il est intéressant d'ajouter dans le plan les quelques essais déjà effectués pour lesquels les résultats sont connus. De plus, on peut aussi intégrer dans le plan, des combinaisons d'essais rencontrés dans la littérature.

III.2.5. DEFINIR LE DOMAINE EXPERIMENTAL

Le domaine expérimental délimite la zone d'étude des facteurs. Il est différent selon le problème de l'expérimentateur (plan de criblage ou plan d'optimisation). Dans les deux cas, il représente les combinaisons d'expériences choisies parmi l'ensemble des combinaisons possibles des facteurs principaux.

Il faudra tenir compte des contraintes précédemment citées pour sa construction. Nous verrons des exemples différents de domaines expérimentaux dans les études de cas (**Partie C**).

Chapitre IV : Analyses et résultats

IV.1. ACQUISITION PROGRESSIVE DES RESULTATS

L'expérimentateur mène les expériences en respectant le plan d'expérimentation. Cette étape est très importante, puisque c'est de la bonne réalisation des expériences que l'on pourra tirer les informations capitales qui seront exploitées par la suite. Il est recommandé d'avancer progressivement et de réorienter voire d'affiner les études en fonction des premiers essais réalisés. Ainsi, une première ébauche permettra de mieux choisir les nouvelles orientations à prendre pour préciser les seuls points intéressants de l'étude et pour abandonner ceux qui ne le sont pas. C'est pour cela qu'il est préconisé de procéder de manière progressive. La première série d'expériences conduit à des conclusions provisoires, ensuite on lance une nouvelle série d'essais. L'expérimentateur accumule ainsi les résultats dont il a besoin et s'arrête dès que le but est atteint. Il est également conseillé de randomiser l'ordre des essais car si le processus subit une dérive, en n'effectuant pas les essais de manière aléatoire, on va amplifier aux effets des facteurs les effets des dérives. Cette **randomisation** consiste à tirer au hasard l'ordre des essais pour supprimer l'influence des facteurs perturbateurs non identifiés pouvant être corrélés avec l'ordre des essais.

- Randomiser et respecter l'ordre des essais, problème d'ajustement, ce qui jouera sur les d_i ,
- Créer dès que possible des répétitions d'expériences, on agit ici sur les g_i ,
- Etre vigilant sur le respect du plan, joue sur les estimateurs des coefficients b_i ,
- Réorienter les études en fonctions des premiers essais

IV.2. ANALYSE GLOBALE DES RESULTATS D'ESSAIS

L'analyse globale des résultats d'essais consiste à analyser de manière générale les résultats du plan d'expérimentations. On peut utiliser des outils d'analyse de données (voir **Partie B**) pour voir l'étendue des réponses obtenues ou pour savoir s'il y a dans les expériences du plan, des expériences atypiques. Il est important de regarder si la solution optimale que l'on désirait pour une réponse n'est pas présente. Dans ce cas, il n'est pas nécessaire de continuer l'analyse puisque l'objectif est atteint.

IV.3. ANALYSE MATHÉMATIQUE

L'estimation d'un modèle mathématique est souvent gérée à l'aide d'un outil informatique (**NemrodW**, **Statgraphics**, **Xlstat**). L'estimation des coefficients des modèles s'effectue dans les exemples de la **Partie C**, à partir de la méthode des moindres carrés. Nous allons en rappeler ici le principe général.

IV.3.1. NOTATIONS MATHÉMATIQUES

Pour analyser les résultats d'un plan d'expériences, il est indispensable de recourir à une notation vectorielle, matricielle et mathématique de façon à simplifier l'analyse des résultats.

Pour une réponse Y , on note :

Y : le vecteur colonne de la réponse expérimentale y

\hat{Y} : le vecteur colonne de la réponse calculée \hat{y}

X : la matrice du modèle. Cette matrice possède N lignes et p colonnes, elle se construit directement à partir de la matrice d'expériences

tX : la transposée matricielle du vecteur X

B : le vecteur colonne des estimateurs des coefficients du modèle b_j

E : le vecteur colonne des résidus e_i .

IV.3.2. MATRICE DU MODELE, MATRICE D'INFORMATION ET DE DISPERSION

La notion de matrice du modèle traduit l'application de l'équation du modèle pour chacune des combinaisons définies à partir de la matrice d'expériences. De manière générale, la matrice du modèle est définie à partir de N lignes et de p colonnes. Le nombre de colonnes correspond au nombre de coefficients à estimer dans le modèle.

La matrice d'information est obtenue à partir du produit matriciel entre la transposée de la matrice du modèle (tX) et la matrice du modèle (X) elle-même. Il s'agit d'une matrice carrée de rang p qui joue un rôle essentiel dans la construction des plans optimaux.

Parmi les critères d'optimalité les plus répandus, celui qui consiste à définir des combinaisons des niveaux des facteurs de manière à maximiser le déterminant de la matrice d'information s'appelle le critère de D-optimalité. En présence d'un arrangement orthogonal, la matrice d'information est diagonale. La méthode des moindres carrés conduit ensuite à inverser la matrice d'information pour définir la matrice de dispersion, les termes diagonaux de cette matrice étant proportionnels à

l'incertitude associée à l'estimation des coefficients du modèle. La qualité de la matrice d'expériences dépend directement de la qualité des coefficients.

La **Figure A.6** résume les différentes étapes mathématiques pour l'obtention de ces matrices.

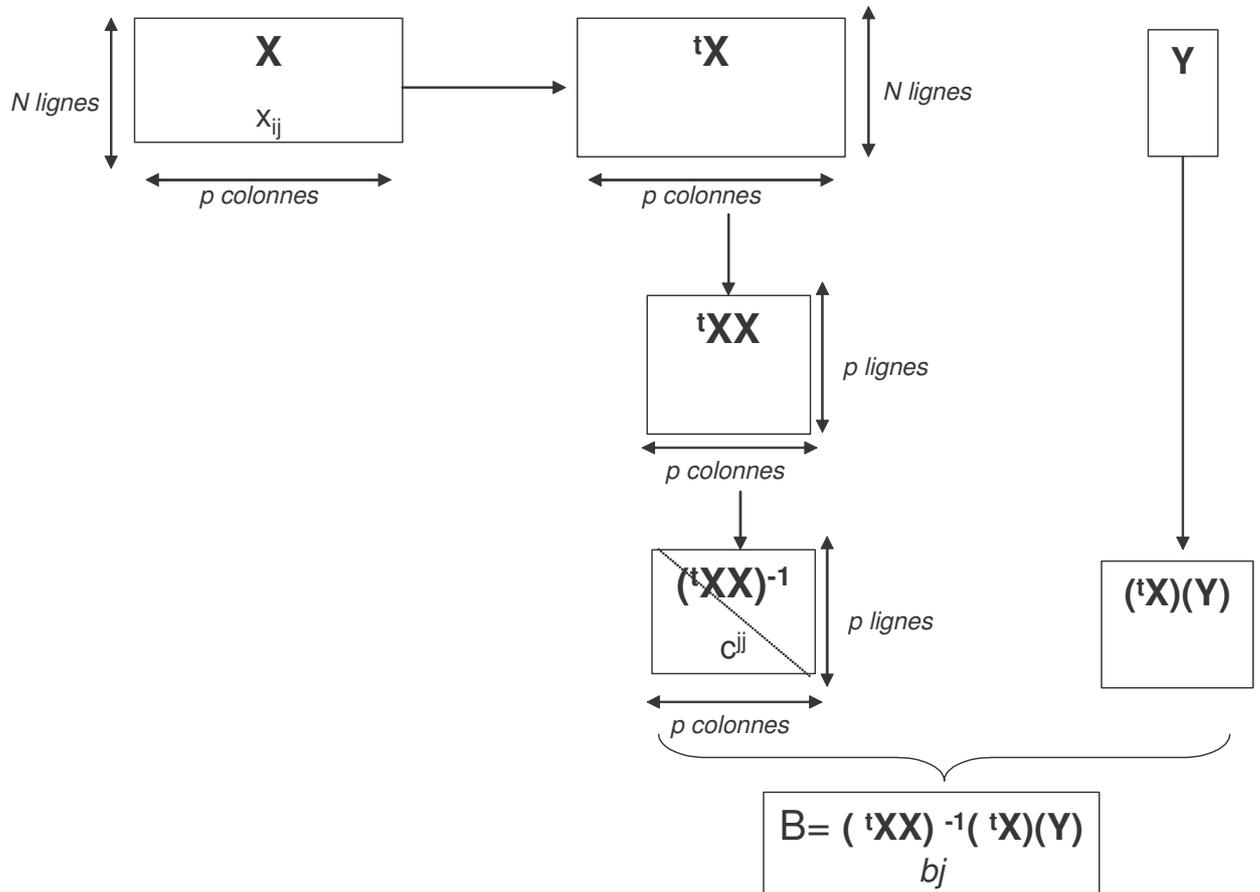


Figure A.6 Illustration des différentes matrices

La qualité des estimateurs des coefficients (b_i) est directement donnée par les termes diagonaux de la matrice de dispersion.

A partir de la relation : $\text{var}(b_i) = c^{ii} \times \sigma^2$ on évalue un critère de qualité de la matrice d'expériences appelé critère A.

Avec σ^2 : estimation de l'erreur expérimentale ou $\text{var}(e_i)$,

c^{ii} : termes diagonaux de la matrice de dispersion $({}^tXX)^{-1}$, appelés coefficients de variance.

La somme de ces éléments diagonaux de cette matrice s'appelle aussi la **trace**. L'objectif d'un plan d'expériences est de minimiser la trace. On pourra également regarder un indicateur appelé l'indice de conditionnement (condition number). L'indice de conditionnement est le rapport des plus grandes et plus petites valeurs singulières de X .

Ce nombre représente une mesure de la sphéricité du plan (orthogonalité). Tous les plans factoriels, sans point au centre ont un indice de conditionnement de 1 et les points du plan sont situés sur la surface d'une sphère. On note que lorsque :

- Condition number < 4 : matrice de qualité optimale
- $4 < \text{Condition number} < 6$: matrice de qualité suffisante
- Condition number > 6 : l'information apportée par la matrice est insuffisante.

Il est évident que la qualité de la matrice d'expériences devra être regardée avant de réaliser les expériences.

IV.3.3. LA REGRESSION MULTILINEAIRE

La **Figure A.7** résume les différentes étapes de cette méthodologie. Les applications dans la **Partie C**, détailleront au mieux ces calculs.

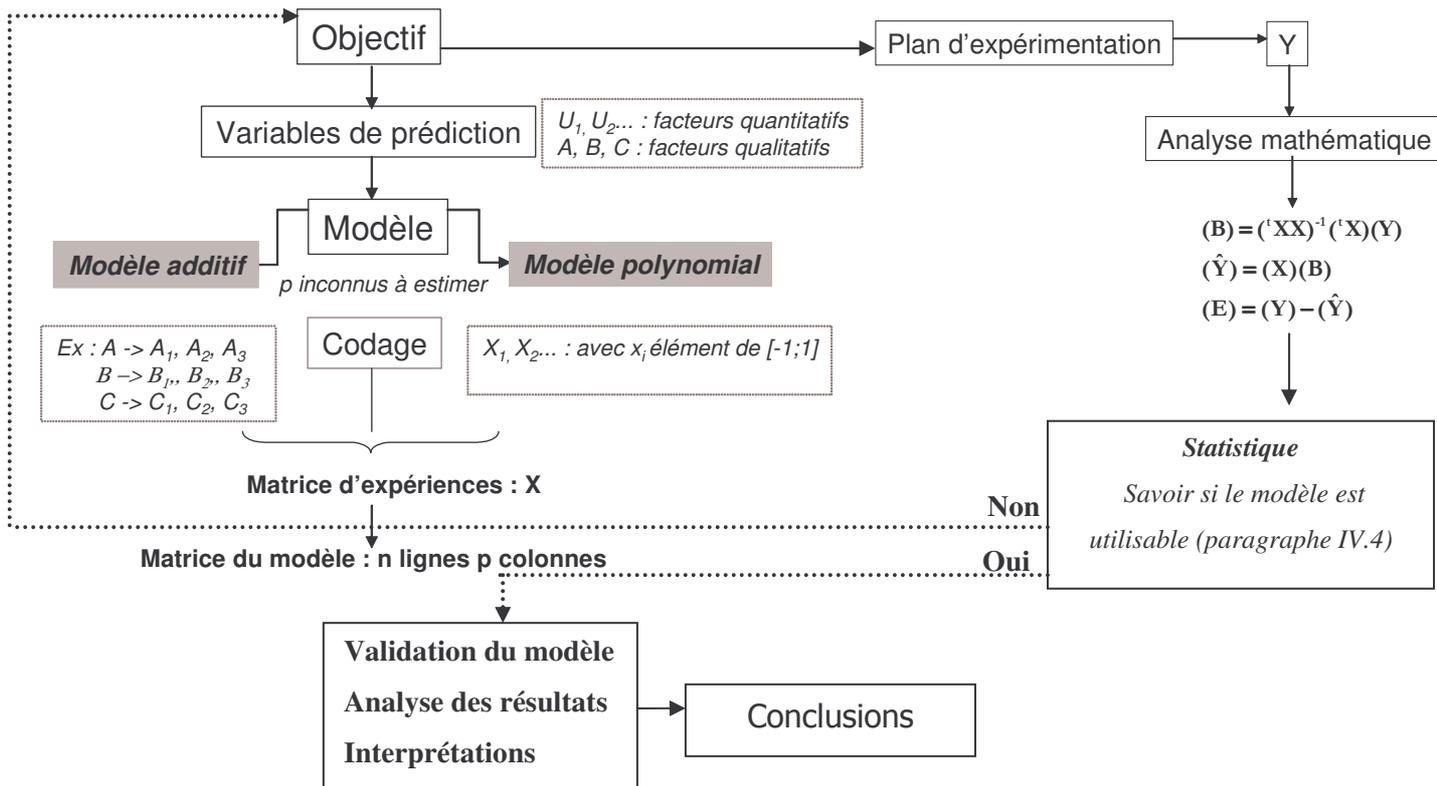


Figure A.7 Méthode des moindres carrés et méthodologie des plans d'expériences

La méthode de régression multilinéaire utilise la méthode des « moindres carrés » qui va nous permettre d'estimer les p coefficients sans biais et les N résidus du modèle empirique. L'estimateur des moindres carrés est celui qui minimise l'écart entre y_i et \hat{y}_i [12]. On cherche les coefficients b_i du vecteur colonne B tels que les estimations de la réponse « ajustent » le mieux possible les observations.

Une fois le plan d'expérimentation réalisé, l'expérimentateur mesure les réponses du système étudié ce qui nous permet de calculer les coefficients b_j du modèle. Il ne reste plus qu'à valider le modèle.

IV.4. ANALYSE STATISTIQUE DES RESULTATS

Toutes les hypothèses faites en amont d'un plan d'expérimentation doivent être vérifiées pour pouvoir valider ensuite les conclusions. C'est à cette étape que le recours aux outils statistiques intervient. Nous relativiserons leur usage et nous nous contenterons de règles pragmatiques utiles à des non-spécialistes de ces méthodes. Toutefois, il est possible de se reporter à la bibliographie [12] pour avoir des informations complètes sur l'usage des statistiques dans l'analyse des plans d'expériences. On peut estimer la qualité descriptive et prédictive des modèles obtenus. Dans tout les cas, l'utilisateur des tests statistiques est étroitement lié à la connaissance de **l'écart type expérimental** σ^2 , estimé par des essais de répétabilité.

Les valeurs expérimentales introduites dans le modèle sont entachées d'erreurs (somme d'erreurs systématiques et aléatoires) qui se transmettent aux coefficients b_j du modèle puis aux valeurs calculées. Des tests statistiques permettent d'évaluer la qualité du modèle (descriptive et prédictive), sa validation (**Analyse de la variance**) et la significativité des coefficients (**test de Student**).

La mise en œuvre de tests statistiques doit permettre aux opérateurs de porter un jugement sur les résultats obtenus à savoir :

- Un modèle décrivant la variation de la réponse dans le domaine expérimental
- Des estimations des coefficients associés aux différents monômes du modèle
- Des résidus traduisant les écarts entre les valeurs mesurées et les valeurs calculées

IV.4.1.a. Qualité descriptive

On peut aussi chercher à savoir si le modèle explique bien l'ensemble des résultats en calculant le

coefficient de détermination R^2 tel que :
$$R^2 = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

Ce coefficient R^2 est aussi appelé coefficient de détermination. Plus R^2 sera grand, meilleur sera le modèle (les réponses calculées seront fortement corrélées avec les réponses expérimentales). Cependant, si le nombre d'expériences est égal au nombre d'inconnues du système, le coefficient R^2

sera toujours égal à 1. C'est pour éviter cela que le coefficient de détermination ajusté (R_a^2) a été introduit. Le R_a^2 ajusté est défini comme la différence à 1 du rapport entre le carré moyen des écarts des résidus et le carré moyen des écarts expérimentaux.

- $R_a^2 = 1 - \frac{N-1}{N-L}(1-R^2)$ avec N le nombre d'essais et L le nombre de coefficients p indépendants pour $N \neq L$. Dans le cas où $N=L$, on dira que le modèle est descriptif.
- De même que pour R^2 , le meilleur modèle à L variables sera obtenu pour R_a^2 maximum.

Ce critère permet d'apprécier la qualité descriptive du modèle et permet de comparer des modèles qui n'ont pas le même nombre de termes. Il ne nous indique pas si le modèle représente bien le phénomène.

IV.4.1.b. Qualité des p coefficients du modèle

Comme nous venons de l'écrire dans le paragraphe IV.2.3, cette estimation se calcule à partir de la relation : $\text{var}(b_j) = c^{jj} \times \sigma^2$. L'écart type de répétabilité représentant l'incertitude sur les valeurs mesurées, il est clair que la qualité des coefficients sera d'autant meilleure que les coefficients de variance c^{jj} seront faibles. La construction d'un plan d'expériences conduit à homogénéiser l'intervalle de confiance

L'incertitude sur les coefficients est alors donnée par la formule :

$$\delta_j = \pm t \sqrt{\text{var}(b_j)}$$

avec t , la valeur du fractile de la loi de **Student** (voir **Annexe 2**). Dans ce cas, on peut écrire l'ensemble des coefficients sous la forme : $b_j \pm \delta_j$.

Connaissant les coefficients du modèle et leurs incertitudes, on calcule la réponse calculée \hat{y}_i qu'il faut comparer à la réponse mesurée y_i .

IV.4.1.c. Fonction de variance de la réponse calculée

Si on prend un point $A(x_1, x_2, x_3)$, on peut à présent grâce aux coefficients, estimer une réponse calculée au point A. Par exemple, si on prend le modèle suivant, on obtient :

$$\hat{y}_{iA} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{13}x_1x_3$$

Dans cette équation, les x_i sont supposés connus sans erreurs, donc l'incertitude de la réponse calculée ne dépend que des coefficients b_i proposés par le modèle.

Ainsi, l'incertitude sur la réponse calculée au point A peut être estimée à partir de la relation :

$$\hat{y}_{iA} = d_A \times \sigma^2 \text{ avec}$$

d_A : fonction de variance, cette fonction dépend des coordonnées de A et de $\text{var}(b_i)$.

La construction d'un plan d'expériences conduit à uniformiser en les minimisant les valeurs de la fonction de variance.

IV.4.1.d. Analyse de la variance

L'analyse de la variance (**ANOVA** ou **AN**alysis **O**f **V**ariance) permet de comparer les variances des valeurs calculées par le modèle et les résidus. Cette analyse constitue un test statistique (**test de Fisher-Snedecor voir annexe 2**) et elle fournit les réponses aux deux questions suivantes

- Est-ce que le modèle nous apporte quelque chose ? La régression est elle significative ? Explique-t-elle en grande partie les variations observées de y_i ? C'est le test de signification globale de la régression que nous appellerons le **test de validation 1**.
- Est-ce que le modèle représente bien le phénomène ? Si oui, le modèle introduit un biais acceptable. Sinon, il convient d'en chercher un autre. La première validation concerne donc l'analyse des résidus du modèle. Nous appellerons cette étape, le **test de validation 2**.

Test de validation 1

Le test de **Fischer-Snedecor** (F_1) va nous dire si le modèle nous apporte quelque chose, si l'équation établit bien une relation entre la variation des facteurs et de la réponse, ou si c'est dû à un changement, une fluctuation aléatoire de la réponse dans le domaine expérimental.

On appelle SCM (somme des carrés moyens) l'expression suivante :

$$SCM = \sum_{i=1}^N (y_i - \bar{y})^2 \text{ avec } \text{ddl}_{SCM} = p-1$$

Or SCT (somme des carrés totale) = SCM+SCRES

$$\text{Avec } SCT = \sum_{i=1}^N (y_i - \bar{y})^2 \text{ avec } \text{ddl}_{SCT} = N-1 \text{ et } SCRES = \sum_{i=1}^N \epsilon_i^2$$

On peut également calculer le coefficient de corrélation à l'aide de ces notations :

$$R^2 = \frac{SCM}{SCT} \text{ et on obtient } R_{\text{ajusté}}^2 = 1 - \frac{SCRES/(N-p)}{SCT/(N-1)} \text{ avec}$$

$$\text{Soit } F_1 = \frac{SCM/(p-1)}{SCRES/(N-1)}$$

Pour qu'il existe le moins de résidus possible, il faut que ce rapport soit le plus grand possible. Dans ce cas on pourra dire que le modèle apporte de l'information.

Le rapport F_1 exprime donc le rapport de deux variances. Pour le $i^{\text{ème}}$ traitement du plan d'expériences, la valeur observée y_i du résultat d'essai représente une valeur particulière de la variable aléatoire caractérisant la réponse pour ce traitement. Donc, la statistique F_1 est elle-même une variable aléatoire dont les valeurs suivent une fonction de répartition théorique, appelée loi de F ou encore loi de Snedecor. Ici l'hypothèse H_0 du test de **Fischer-Snedecor** est la suivante : *le modèle ne permet pas de décrire la variation des résultats d'essais.*

On utilise cette loi pour savoir à partir de quelle valeur particulière, appelée valeur critique, le numérateur de la quantité F_1 est significativement supérieur au dénominateur.

En d'autres termes, F_1 nous renseigne sur la probabilité **ProbF₁** de rejeter l'hypothèse nulle. Donc dans ce cas, nous voulons obtenir la plus petite valeur possible de cette probabilité. On dira qu'au dessous de 5%, nous n'avons pas de raison de rejeter le modèle.

Le **Tableau A.8** résume les différents calculs et notations de ce test.

Source	Somme des carrés	ddl		Test F	Probabilité
Modèle	SCM	p-1	SCM/p-1	F_1	ProbF ₁
Écarts	SCRES	N-p	SCRES/N-p		
Total	SCT	N-1			

Tableau A.8 Récapitulatif du Test 1

Toutes ces informations sont données par les logiciels cités dans un tableau que l'on appelle le **tableau d'analyse de régression** (voir **Partie C**).

Test de validation 2

Bien que les écarts entre les valeurs observées et les valeurs calculées par le modèle aient été minimisés par le choix de la méthode des moindres carrés, il faut s'assurer que localement, les résidus ne soient pas anormalement importants. Tout d'abord on calcule à partir de l'équation du modèle les différents y_i .

On obtient alors la valeur des résidus : $\epsilon_i = y_i - \hat{y}_i$

On calcule ensuite la somme des carrés des résidus notée **SCRES** : $SCRES = \sum_{i=1}^N \epsilon_i^2$

Or $SCRES = SCE1 + SCE2$

avec SCE1 l'erreur due au manque d'ajustement du modèle, appelé aussi le biais ou encore le Lack of Fit,

SCE2 l'erreur pure ou naturelle encore appelé Pure Error.

Pour estimer si ce que l'on obtient est acceptable, on procède à un **test de Fisher-Snedecor** ou **loi de F**. Or, la loi de F dépend du nombre de **degrés de liberté (ddl)** de chacune des deux populations

donc :

$$F_2 = \frac{SCE1/ddl_{SCE1}}{SCE2/ddl_{SCE2}}$$

Rappel sur le **ddl** : le nombre de degré de liberté du modèle est la somme des actions qu'il comporte. Le degré de liberté d'une action représente le nombre de coefficients indépendants que la régression permettra d'estimer [12].

On a donc : $ddl_{SCRES} = ddl_{SCE1} + ddl_{SCE2} = N - p$

Le test F (F_2) va être calculé en posant une hypothèse nulle (H_0 : *Hypothèse nulle que le rapport soit égal à 1, c'est-à-dire si les résidus liés au modèle sont du même ordre de grandeur que l'erreur naturelle, on teste si le modèle est juste et fidèle*). On obtiendra la probabilité (**ProbF₂**) de trouver un rapport égal à 1. On dira qu'au dessus de 5%, nous n'avons pas de raison de rejeter le modèle.

ProbF₂ > 5% ⇒ Pas de raison de rejeter le

Ce test n'est faisable qu'en présence de répétitions ce qui permet de connaître SCE2.

Le **Tableau A.9** résume les différents calculs et notations de ce test.

Source	Somme des carrées	ddl		Test F	Probabilité
Lack of Fit	SCE1	ddl _{SCE1}	SCE1/ddl _{SCE1}	F ₂	ProbF ₂
Pure Error	SCE2	ddl _{SCE2}	SCE2/ddl _{SCE2}		
Résidus	SCRES	ddl _{SCRES}			

Tableau A.9 Récapitulatif du Test 2

Il faut faire attention aux conclusions hâtives ! En effet :

- ⇒ si on a une « Erreur » très faible et une mauvaise « validité », on peut aussi avoir une bonne régression,
- ⇒ si on a une « Erreur » élevée et une bonne « validité », on peut aussi avoir une mauvaise régression.

Récapitulatif :

Meilleur est l'ajustement du modèle si :

- R² proche de 1
- R_a² maximum
- ProbF₁ < 5% F₂ le plus petit possible
- ProbF₂ > 5% F₁ le plus grand possible

IV.5. ANALYSE GRAPHIQUE DES RESULTATS

Un des grands avantages des plans d'expériences est la présentation des résultats sous forme graphique (voir figure ci-dessous)

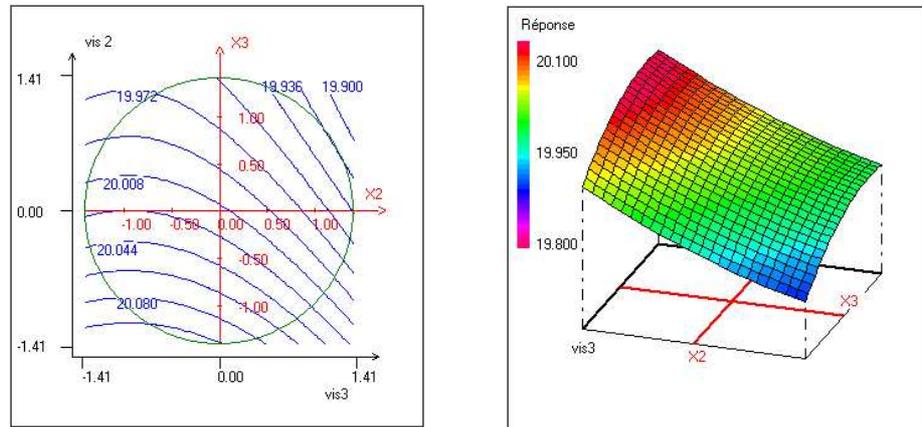


Figure A.8 Courbe d'iso-réponses et surface de réponses

Ce sera également le cas pour les méthodes d'analyses de données développées dans la **Partie B**. Pour chacune des méthodes, nous obtenons des graphiques différents qui répondent aux questions posées.

Les graphiques sont avant tout un outil d'aide à l'interprétation des résultats mais, ils permettent également de manière plus communicative lors d'une réunion de tirer plus rapidement des conclusions et ainsi d'orienter la poursuite d'une étude.

IV.6. VALIDATION DES HYPOTHESES ET RECHERCHE DE SOLUTIONS AU PROBLEME

L'analyse des résultats d'essais permet d'identifier une combinaison optimale des facteurs qui n'a pas forcément fait l'objet d'un essai dans le plan. Il faudra alors tester la combinaison optimale, qui n'a pas été faite dans le plan d'expériences, ce qui arrive fréquemment. Avec cet essai final, on rejettera ou pas le modèle proposé.

De plus, quand on arrive à l'étape de l'analyse finale, il faut savoir si les résultats obtenus par l'analyse sont plausibles ou non.

Dans l'essai de validation, soit l'essai confirme les informations

- si oui, le plan d'expériences aura joué son rôle et permis de mettre en évidence les facteurs influents et/ou d'optimiser la réponse.

- si non, il faudra alors examiner les conditions dans lesquelles ont été effectuées les expériences et vérifier si des facteurs influents n'ont pas été oubliés ou s'il ne se cache pas un effet d'interaction entre deux facteurs ou encore d'autres choses
- A ce stade, il faut décider en collaboration avec l'expérimentateur quelles sont les nouvelles stratégies à entreprendre. Les conclusions obtenues seront capitalisées dans des rapports de synthèse dans lesquels devront apparaître les influences mises en évidence, la solution optimale retenue et les perspectives.

Conclusion

Le gain de temps, de coût et de main d'œuvre fait de la méthodologie des plans d'expérience un outil incontestablement efficace pour mettre en œuvre des expériences.

Nous avons vu à travers cette partie, les différentes articulations méthodologiques des plans d'expériences. On retiendra surtout les grandes étapes de la fiche méthodologique (**Figure A.9**).

La méthodologie des plans d'expériences ne nous donne pas d'informations sur les corrélations des réponses entre elles. Mais, chaque plan d'expériences générant des données, il sera donc possible d'utiliser ensuite des méthodes d'analyses multidimensionnelles des données (voir **Partie B**) telles que l'analyse en composantes principales, méthodes visant à obtenir des graphiques traduisant les corrélations entre les variables, qu'il s'agisse de facteurs ou de réponses.

Nous verrons donc à travers les études de cas de la **Partie C** de ce manuscrit, l'utilisation croisée de ces différentes méthodes.

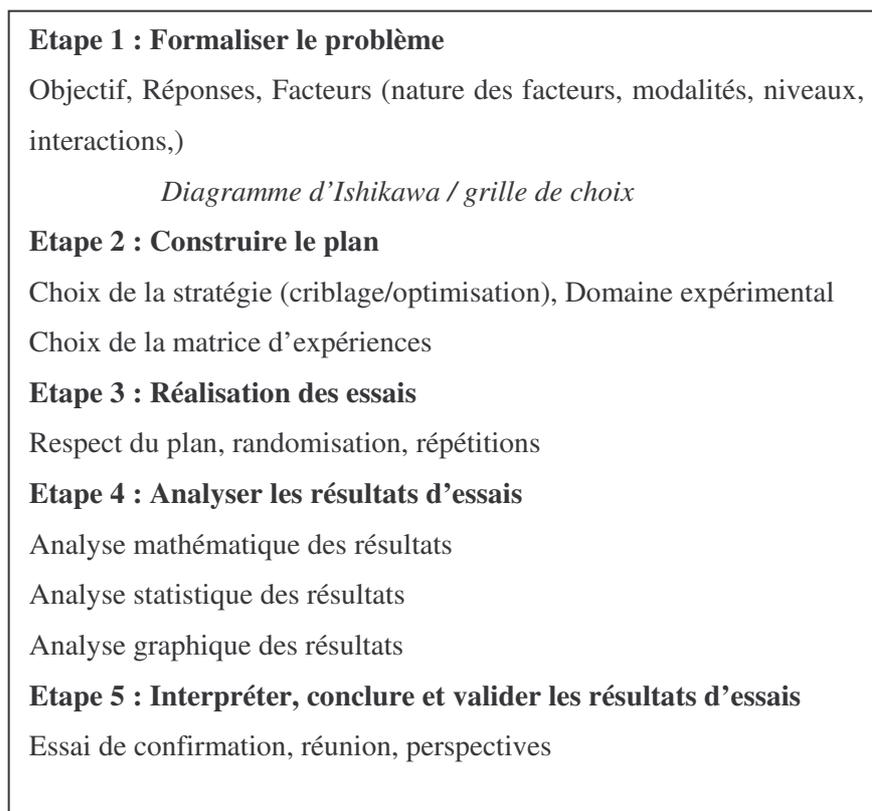


Figure A.9 Fiche méthodologique

Rappel des notations

<p>K : nombre de facteurs maîtrisés ou facteurs principaux</p> <p>s_i : nombre de modalités d'un facteur</p> <p>U_i: niveau d'un facteur quantitatif pris dans un intervalle [U_{mini} ; U_{maxi}]</p> <p>m_i : modalités d'un facteur qualitatif</p> <p>N : nombre d'expériences du plan</p> <p>N_{total}: nombre des combinaisons possibles des facteurs (plan factoriel complet)</p> <p>N_{total} = s₁^{k1} s₂^{k2} ... s_i^{ki} (Matrice asymétrique)</p> <p>N_{total} = s^k (Matrice symétrique)</p> <p>Y_j: j^{ème} réponse du système</p> <p>X_j: j^{ème} facteur d'entrée</p> <p>y_i: réponse expérimentale</p> <p>η : résultat théorique, générique</p> <p>ŷ_i : réponse calculée</p> <p>e_i : erreur résiduelle</p> <p>b_j: coefficients du modèle</p> <p>p : nombre de coefficients du modèle</p> <p>σ_{exp} : écart type expérimental</p> <p>c^{jj}: coefficients de variance</p>	<p>Y : vecteur colonne de y_i</p> <p>^tY : transposée matricielle du vecteur Y</p> <p>η : vecteur colonne de η_i</p> <p>X : matrice du modèle</p> <p>B : vecteur colonne des b_i</p> <p>E : vecteur colonne e_i</p> <p>(^tXX)⁻¹ : matrice de dispersion</p> <p>^tXX : matrice d'information</p> <p>N₀: nombre de répétitions au centre</p> <p>F₁ : Rapport du Test Fisher 1</p> <p>F₂ : Rapport du Test Fisher 2</p> <p>ddl : degré de liberté</p> <p>R² : coefficient de détermination, ou carré du coefficient de corrélation</p> <p>SCRES : somme des carrés des résidus</p> <p>R_a² coefficient de détermination ajusté</p> <p>var(η_{ia}) = f(ρ) : précision de η_i sur un cercle de rayon ρ</p> <p>d_A : fonction de variance</p> <p>δ_i : incertitude sur les coefficients</p>
---	---

Bibliographie Partie A

- [1] G. SADO, M. C. SADO, Les plans d'expériences, de l'expérimentation à l'assurance qualité, Collection AFNOR, 1991.
- [2] P. SCHIMMERLING, J.C. SISSON, A. ZAIDI, *Pratique des plans d'expériences*, Technique & documentation, 1998.
- [3] R.A. FISHER, *The Design of Experiments*, Oliver and Boyd, Ed. Edimbourg 5^{ème} édition, 1949.
- [4] W.G COCHRAN, G.M. COX, *Experimental design*, 2^e edition Wiley, 1957
- [5] YATES, *The Design and Analysis of Factorial Experiments*, Bulletin 35, Imperial Bureau of Soil Sciences, Harpenden Macmillan, 1937.
- [6] R.L PLACKETT, J.P. BURMAN, *The design of optimum multifactorial experiments*, Biometrika n°3, 1946.
- [7] G.E.P. BOX, W.G. HUNTER, J.S. HUNTER, *Statistics for Experimenters, An introduction to design, Data analysis, and Model Building*, Wiley, Ed. New York, 1978.
- [8] G. TAGUCHI, *Introduction to quality Engineering*, Asian Productivity Organisation. American Supplier Institute Inc. Unipub-Kraus, 1986.
- [9] G.TAGUCHI, S. KONISHI, *Taguchi Methods- Orthogonal Arrays and Linear Graphs*, American Supplier Institute Inc. Dearborn, U.S.A. 1987.
- [10] J. DEMONSANT, *Comprendre et mener des plans d'expériences*, Ed. AFNOR, 1996.
- [11] J. ALEXIS, P. ALEXIS, *Pratique industrielle des plans d'expériences – La qualité à moindre coût : l'approche Taguchi*, Afnor, Ed. Paris, 1999.
- [12] D. BENOIST, Y. TOURBIER, S. GERMAIN-TOURBIER, *Plans d'expériences : construction et analyse*, Technique & documentation- Lavoisier, Ed.Paris, 1994.
- [13] T. LUNDSTEDT, E. SEIFERT, L. ABRAMO, B. THELIN, A. NYSTROM, J. PETTERSEN, R. BERGMAN, *Experimental design and optimization*, Chemometrics and intelligent laboratory systems 42, 1998.
- [14] H. SCHEFFE, *The Analysis of Variance*, Wiley & Sons, Ed. New York.

Partie B :

«Théorie sur l'analyse de données»

Introduction

I.1. ETAT DE L'ART

Aujourd'hui, l'analyse de données est couramment utilisée dans bon nombre d'activités ou de professions : économistes, médecins, politiciens, chefs d'entreprises, ingénieurs et bien des disciplines de bases comme : la physique, la chimie, la géologie, l'économie, la sociologie, la psychologie [1][2], la biométrie, les sciences médicales. Toutes utilisent des « faits numériques » qu'on appelle **données**. Mais, ces techniques ne sont pas encore très répandues dans le domaine de l'électronique et notamment dans toutes les disciplines relevant de la technologie comme les techniques d'élaboration des dépôts où pourtant elles semblent être un outil pertinent. Nous verrons à travers les applications traitées dans la **Partie C** de ce manuscrit, que l'analyse des données nous a été d'une grande utilité, dans l'aide à l'interprétation de résultats ou dans la prédiction d'expériences.

I.2. HISTORIQUE ET DEFINITION

L'analyse de données désigne l'ensemble des méthodes à partir desquelles on collecte, organise, résume, présente, étudie des données pour permettre d'en tirer des informations. Définie ainsi, l'analyse de données n'est pas une science nouvelle. Déjà dans l'Egypte ancienne, le Pharaon l'utilisait pour établir le montant de l'impôt basé sur la hauteur de crue du fleuve donnant ainsi une estimation des récoltes ! [3][4]. **H. Hotelling** dans les années 30 posait les fondements de l'analyse en composantes principales [1] et de l'analyse canonique [5] en développant les travaux de **C. Spearman** [2] et de **K. Pearson** [6] qui dataient du début du siècle. Jusqu'aux années 60, ces méthodes étaient perfectionnées et s'enrichissaient de variantes mais toutes restaient inabordables pour les praticiens car elles nécessitaient une masse considérable de calculs. Aujourd'hui, démocratisées grâce à l'outil informatique, ces méthodes sont de plus en plus utilisées.

I.3. UTILISATION DE L'ANALYSE DE DONNEES

Avant toute analyse statistique, il est indispensable de consacrer du temps à examiner les données d'un point de vue général. En tout premier lieu, il est important de connaître le langage utilisé pour décrire ces données et d'en faire un tableau facilement exploitable. On pourra immédiatement détecter un phénomène inattendu, une erreur de mesure ou une simple faute de frappe. Une connaissance générale des données permettra de distinguer les individus actifs dans l'analyse, des individus d'illustration. C'est dans cette optique, que nous avons choisi dans le **chapitre II** de cette partie de décrire quelques-unes de ces analyses dites unidimensionnelles et bidimensionnelles.

La statistique classique s'est axée sur l'étude d'un nombre restreint de paramètres mesurés sur un petit ensemble d'individus. Elle a développé les notions d'estimations et de tests fondés sur des hypothèses probabilistes très restrictives. Cependant dans la pratique, les individus observés sont fréquemment décrits par un grand nombre de caractères. Les méthodes d'analyse de données permettent une étude globale des individus et des variables en utilisant généralement des représentations graphiques suggestives. Les données peuvent être analysées selon plusieurs points de vue. La recherche des ressemblances ou des différences entre individus peut être un objet de l'analyse. On considère que deux individus se ressemblent lorsque leurs profils selon les différents paramètres sont voisins. En utilisant des méthodes factorielles (comme par exemple l'Analyse en Composante Principale, voir **chapitre III**), il est possible de représenter ces proximités entre individus sur un même graphique. Les méthodes de classification (voir **chapitre IV**) permettent de regrouper les individus en catégories homogènes. La description des relations entre les paramètres peut être un autre objet de l'analyse : deux paramètres sont considérés comme liés ou corrélés s'ils varient de la même façon sur les différents individus [7].

Selon le type de problème et la nature des données, on choisit la méthode appropriée. Cette partie sur l'analyse de données abordera plus particulièrement trois méthodes considérées comme fondamentales dans l'analyse de données : l'analyse en composante principale (**chapitre III**), la classification ascendante hiérarchique (**chapitre IV**) et l'analyse factorielle discriminante (**chapitre V**).

De manière générale, ces techniques nous sont apparues comme pertinentes pour décrire un ensemble d'expériences, caractériser de manière plus visuelle leur identité, détecter des paramètres d'expérimentations importants ou suspects et surtout de relier ces paramètres entre eux.

Chapitre II Pré-traitement des données

Avant de commencer une analyse de données multidimensionnelles, il est nécessaire et indispensable de faire ce qu'on appelle ici un pré-traitement des données. La première chose est de connaître un « minimum » le langage de ces méthodes ainsi que les bases sur certaines notions mathématiques et statistiques. Ce chapitre a donc pour objet de rappeler, de manière simplifiée, l'ensemble de ces notions pour pouvoir comprendre au mieux ces méthodes, mais également de préciser quelques concepts fondamentaux sur la nature des données.

II.1. LA NATURE DES DONNEES [3]

II.1.1. DEFINITION DES VARIABLES

Les variables correspondent aux intitulés des colonnes du tableau de données, elles sont également appelées paramètres ou caractères ; elles forment un espace dans lequel chaque individu est représenté par un point. Les coordonnées des points sont égales aux valeurs respectives des variables. Il est naturellement impossible de visualiser le nuage de points dès que le **nombre de variables**, noté généralement **p** est supérieur à 3.

II.1.2. DEFINITION DES INDIVIDUS

L'individu est l'entité de base sur laquelle l'observateur réalise un certain nombre de mesures. Il est parfois appelé critère. Chaque **individu** est défini à partir de plusieurs **variables** quantitatives et/ou qualitatives. Une variable est quantitative lorsqu'elle prend ses valeurs sur une échelle numérique, plus précisément lorsque l'ensemble des valeurs qu'elle prend sur les individus est inclus dans l'ensemble des nombres réels et que l'on peut effectuer sur cette variable des opérations algébriques habituelles : additions, multiplications, etc. Elle est qualitative lorsqu'elle prend des modalités non numériques : appréciations, couleurs, descriptions, etc. Les individus constituent une population, on note **n** le **nombre d'individus**.

II.1.3. NOTATIONS

Certains auteurs s'accordent à dire que, dans le cadre de certaines analyses multidimensionnelles de données, les nombres d'individus **n** et de variables **p** doivent être respectivement supérieures à 15 et à 4 afin de ne pas fausser l'analyse. L'élément courant du tableau brut est noté x_{ij} , il correspond à la valeur mesurée ou observée de la $j^{\text{ème}}$ variable pour le $i^{\text{ème}}$ individu. Le tableau de données comprend ainsi **p colonnes et n lignes**. Le **Tableau B.1** résume de façon générale ces différentes notations et le **Tableau B.2** montre un exemple de données avec 5 variables et 35 individus.

Tableau B.1 Modèle de tableau de données à plusieurs variables appelés tableau Individus-Variables

p variables

	X_1	X_j	...	X_p	
n individus	1	X_{11}	X_{1j}	...	X_{1p}
	2	X_{21}	X_{2j}	...	X_{2p}
	3	X_{31}	X_{3j}	...	X_{3p}

	i	X_{i1}	X_{ij}	...	X_{ip}

	n	X_{n1}	X_{nj}	...	X_{np}

Tableau B.2 Exemple de tableau de données brutes

Individus	x1	x2	x3	x4	x5
1	9,5	243	7,3	110,5	218
2	1,0	250	25,9	125,1	2
3	7,2	467	13,9	70,3	8
4	3,2	34	23,5	121,5	19
5	7,2	34	15,7	93,8	263
6	3,2	465	23,9	92,0	1
8	3,0	178	21,7	112,8	11
9	8,0	178	20,1	114,5	316
10	5,2	394	19,9	122,5	12
13	8,2	320	11,9	52,3	36
14	3,0	178	30,0	128,5	2
15	7,8	178	22,4	168,0	108
16	5,2	394	27,5	95,8	2
18	3,1	330	18,4	82,3	1
19	5,7	110	12,6	79,5	11
20	5,5	250	15,7	138,5	131
21	5,2	250	9,3	99,0	3,5
22	1,0	0	85,0	370,0	22
23	2,0	50	70,0	300,0	13
24	4,0	100	49,2	237,0	16
26	1,0	50	52,3	216,0	10
28	4,0	200	37,1	199,0	17
29	10,0	100	30,0	164,0	329
31	2,0	200	30,7	140,0	0
32	4,0	0	26,1	158,0	304
33	10,0	50	25,3	224,0	594
34	1,0	200	40,2	142,0	2
35	2,0	100	19,6	139,0	22
37	10,0	0	17,1	90,0	423
38	10,0	100	21,8	225,0	1100

II.1.4. CODAGE DU TABLEAU DE DONNEES BRUTES [2]

On appelle tableau de données recodées, un nouveau tableau établi à partir de données brutes (voir **Tableau B.2**) au moyen de transformations mathématiques qu'on appelle «recodage»⁽¹⁾.

II.1.4.a. Pourquoi coder les données ?

En présence de grandeurs exprimées dans des unités différentes ou bien présentant des étendues différentes, il peut être délicat d'établir des similitudes à la fois entre variables et entre individus. Les résultats d'une analyse de données sont parfois influencés par l'ordre de grandeur des variables, aussi pour y remédier, il est recommandé de donner la même importance (le même « poids ») aux différentes variables en les réduisant et en les centrant. Ainsi, aucune variable ne pourra dominer une autre variable à cause de ses valeurs numériques. On peut voir sur la **Figure B.1** de façon plus implicite que coder les données revient à « étirer » ou à « compresser » les variables autour d'une même moyenne.

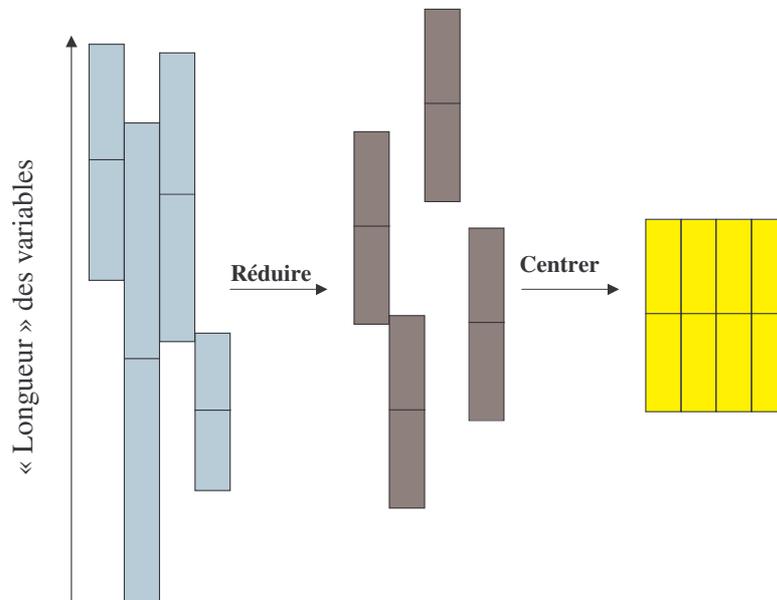


Figure B.1 L'opération de codage (centrage-réduction)

II.1.4.b. Centrage et réduction des données brutes

Dans ce paragraphe, nous présentons un exemple de codage très fréquemment utilisé dans le cadre de l'analyse de données : le codage des données brutes en données centrées et centrées réduites. Pour obtenir des données centrées, il faut remplacer x_{ij} par $(x_{ij} - \bar{x}_j)$ où \bar{x}_j représente la moyenne arithmétique des valeurs prises par la variable x_j .

⁽¹⁾ François LOUVET, *Analyse en composantes principales, Notes de cours, ENSCI, 2000.*

Pour obtenir des données centrées réduites notées X_{ij} , on divise par l'écart type des valeurs prises par la variable x_j (voir paragraphe **A.II.2.1.a.** pour l'écriture mathématique de l'écart type $\sigma(x_j)$ et de la moyenne arithmétique \bar{x}_j).

Ce qui donne :

$$X_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma(x_j)}$$

À titre d'exemple, le **Tableau B.3** représente les données du **Tableau B.2** centrées réduites ; X_1 représente la variable x_1 centrée et réduite, X_2 pour x_2 , et ainsi de suite. On peut par exemple remarquer que les valeurs initiales prises par la variable x_1 (comprise entre 1 et 10) et celles correspondant à la variable x_4 (52,3 à 370) sont remplacées par les variables centrées réduites X_1 et X_4 qui varient entre -1,358 et 1,643 pour X_1 et de -1,363 à 3,207 pour X_4 .

Tableau B.3 Données centrées réduites.

Individus	X1	X2	X3	X4	X5
1	1,380	0,470	-1,188	-0,525	0,360
2	-1,451	0,522	-0,094	-0,315	-0,559
3	0,614	2,142	-0,800	-1,104	-0,533
4	-0,718	-1,091	-0,236	-0,367	-0,486
5	0,614	-1,091	-0,692	-0,766	0,551
6	-0,718	2,127	-0,208	-0,791	-0,561
8	-0,785	-0,016	-0,340	-0,493	-0,520
9	0,880	-0,016	-0,433	-0,468	0,778
10	-0,052	1,597	-0,446	-0,353	-0,516
13	0,947	1,044	-0,918	-1,363	-0,414
14	-0,785	-0,016	0,149	-0,266	-0,557
15	0,814	-0,016	-0,297	0,302	-0,106
16	-0,052	1,597	0,002	-0,737	-0,558
18	-0,752	1,119	-0,535	-0,931	-0,561
19	0,114	-0,523	-0,879	-0,971	-0,521
20	0,048	0,522	-0,693	-0,123	-0,010
21	-0,052	0,522	-1,072	-0,691	-0,552
22	1,380	-1,345	3,391	3,207	-0,472
23	-1,118	-0,971	2,507	2,200	-0,510
24	-0,452	-0,598	1,280	1,294	-0,498
26	-1,451	-0,971	1,462	0,992	-0,526
28	-0,452	0,149	0,567	0,747	-0,494
29	1,547	-0,598	0,150	0,244	0,834
31	-1,118	0,149	0,193	-0,101	-0,565
32	-0,452	-1,345	-0,083	0,158	0,726
33	1,547	-0,971	-0,129	1,107	1,959
34	-1,451	0,149	0,747	-0,072	-0,559
35	-1,118	-0,598	-0,465	-0,115	-0,473
37	1,547	-1,345	-0,609	-0,820	1,232
38	1,547	-0,598	-0,332	1,121	4,111
Variance	1	1	1	1	1
Moyenne	0	0	0	0	0

II.2. STATISTIQUE ELEMENTAIRE

Avant de mettre en oeuvre les outils d'analyse multidimensionnelle des données, il est important d'avoir un résumé d'informations sur chacune des variables du tableau de données, pour lesquelles il est recommandé d'établir quelques statistiques descriptives de base dites unidimensionnelles et bidimensionnelles. Une excellente analyse effectuée sur des données fausses n'a aucune valeur. Toute analyse suppose deux opérations successives : d'abord le recueil des données et ensuite leur traitement. Il est en effet clair que c'est la première de ces deux opérations qui est la plus importante. Dans ce paragraphe, on fera un rappel sur les statistiques de base, indispensable à la compréhension de cette partie, comme par exemple les paramètres de dispersion ou encore la notion de corrélation. Mais surtout, nous détaillerons l'interprétation des outils graphiques servant à déceler des valeurs suspectes (dus par exemple, à des erreurs de saisie des données ou d'expérimentation). Ces représentations seront très utilisées dans la suite de ce manuscrit.

II.2.1. ANALYSE UNIDIMENSIONNELLE DE DONNEES

L'analyse unidimensionnelle de données est l'ensemble des statistiques décrivant les caractéristiques d'une seule variable. L'analyse unidimensionnelle qui nous est apparue la plus pertinente est la boîte à moustache. Pour comprendre et interpréter sa construction, nous allons tout d'abord rappeler quelques définitions.

II.2.1.a. Paramètres de dispersion [8][9]

Les paramètres de dispersion sont des nombres qui mesurent la dispersion (ou la répartition) des individus autour d'un paramètre de position (par exemple la moyenne arithmétique). Ils s'expriment dans la même unité que les données et permettent de comparer des séries statistiques de même nature.

Caractéristique de tendance centrale

Pour chacune des valeurs X_{ij} des variables X_j , on définit la **moyenne arithmétique** des valeurs observées ou mesurées à partir de la formule :

$$\bar{X}_j = \frac{1}{n} \sum_{i=1}^n X_{ij} .$$

Caractéristique de dispersion

La notion de dispersion complète les informations sur la tendance centrale. Il arrive que l'on traduise la dispersion des données à partir de l'**étendue** : $\text{Étendue} = X_{ij_{\max}} - X_{ij_{\min}}$

Variance et écart-type

□ **VARIANCE :**

La **variance** est un indicateur de dispersion reflétant l'importance des fluctuations des valeurs prise par une variable X_j de la série autour de leur moyenne arithmétique. Elle s'écrit : $\text{var}(X_j) = \frac{1}{n} \sum_{i=1}^n (X_{ij} - \overline{X_j})^2$

Notons que, pour des données centrées réduites, la **variance de chaque variable est égale à 1** (**Tableau B.3**). C'est une information qui nous sera nécessaire pour comprendre certaines notions de l'analyse multidimensionnelle de données.

□ **ECART-TYPE :**

Le calcul de l'étendue ne prend en compte que deux valeurs observées ; de plus il s'agit des valeurs extrêmes. On préférera caractériser la dispersion à partir de **l'écart type** expérimental $\sigma(X_j)$:

$$\sigma(X_j) = \sqrt{\text{var}(X_j)}$$

Intervalles et écarts interquantiles

□ **INTERVALLES INTERQUANTILES :**

En statistique, pour toute série numérique de données à valeurs dans un intervalle I , on définit la fonction quantile Q , de $[0,1]$ dans I , par :

$$Q(u) = \inf \{x, F(x) \geq u\}$$

où $F(x)$ est la fonction de répartition qui à un réel x associe la probabilité cumulée $F(x)$ qui désigne la fréquence des éléments de la série inférieurs ou égaux à x . Un intervalle interquantile d'ordre a , I_a , est un intervalle numérique qui contient a % des valeurs observées centrales, en laissant $\frac{(100-a)}{2}$ % des valeurs les plus petites et $\frac{(100+a)}{2}$ % des plus grandes. On note I_a l'intervalle interquantile d'ordre a tel

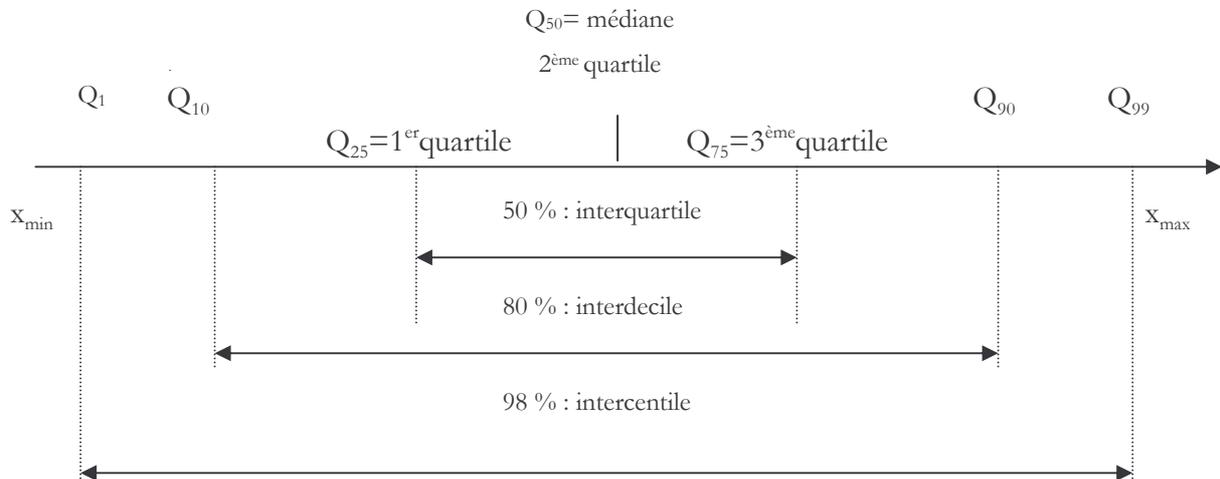
que : $I_a = \left[Q_{\frac{100-a}{2}} ; Q_{\frac{100+a}{2}} \right]$. A titre d'exemple, la **Figure B.2** résume les principaux intervalles

interquantiles. Le nom de l'intervalle interquantile change suivant le pourcentage d'individus contenu dans l'intervalle. Le plus important à connaître est l'intervalle inter*quartile* $I_{50} = [Q_{25} ; Q_{75}]$ qui contient 50 % des observations centrales (individus qui se situent autour de la moyenne arithmétique). On verra au paragraphe II.2.2., l'intérêt de l'intervalle interquartile pour fabriquer des boîtes à moustaches.

Sont également reportés sur la **Figure B.2**, l'intervalle inter*décile* $I_{80} = [Q_{10} ; Q_{90}]$ qui contient 80% des observations centrales et l'intervalle inter*centile* $I_{98} = [Q_1 ; Q_{99}]$ qui en contient 98 %.

□ ECARTS INTERQUANTILES :

On appelle écart interquartile le nombre $Q_{75} - Q_{25}$, écart interdécile le nombre $Q_{90} - Q_{10}$, et écart intercentile le nombre $Q_{99} - Q_1$.



Tous ces paramètres vont nous permettre de construire des boîtes à moustaches.

II.2.1.b. La boîte à pattes ou boîte à moustaches ⁽²⁾

La boîte à pattes ou boîte à moustache (traduction française de l'anglais **Box & Whiskers Plot** ou plus généralement **Box Plot**) recouvre une grande variété de diagrammes en forme de boîtes qui se différencient par leurs constructions, leurs interprétations et leurs usages. La boîte à pattes de **Tuckey** est la plus utilisée pour prendre « connaissance » des données, c'est à dire pour voir au mieux de manière graphique la répartition de la valeur numérique des individus pour une variable choisie. Sur la **Figure B.3** sont reportés les paramètres nécessaires à la construction d'une boîte à pattes. Ce sont :

- la valeur du premier quartile Q_{25} (de la frontière basse jusqu'à Q_{25} il y a 25% des individus donc 25% des effectifs de la variable) qui représente le trait inférieur de la boîte,
- la valeur de la médiane correspondant au deuxième quartile Q_{50} (50% des effectifs) représentée par un trait horizontal à l'intérieur de la boîte,
- la valeur du troisième quartile Q_{75} (75% des effectifs) représenté par le trait supérieur de la boîte,
- les 2 « pattes » inférieure et supérieure, qui sont déterminées à partir de ces équations :

pour la frontière basse : $Q_{25} - 1.5 \times (Q_{75} - Q_{25})$; pour la frontière haute : $Q_{75} + 1.5 \times (Q_{75} - Q_{25})$.

⁽²⁾ Monique LE GUEN, *La Boîte à Pattes de TUCKEY – MATISSE CNRS UMR8595 PARIS*.

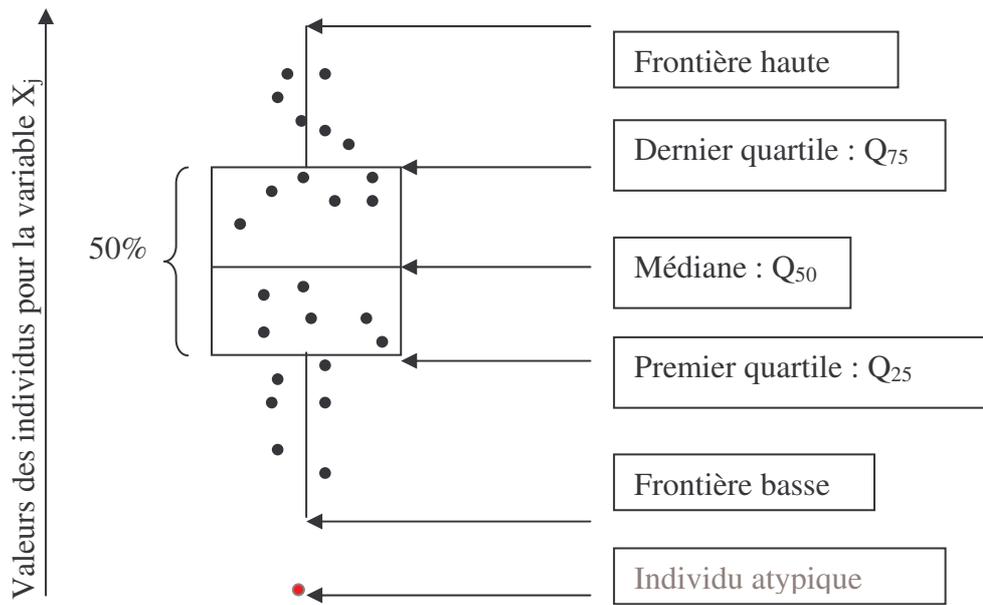
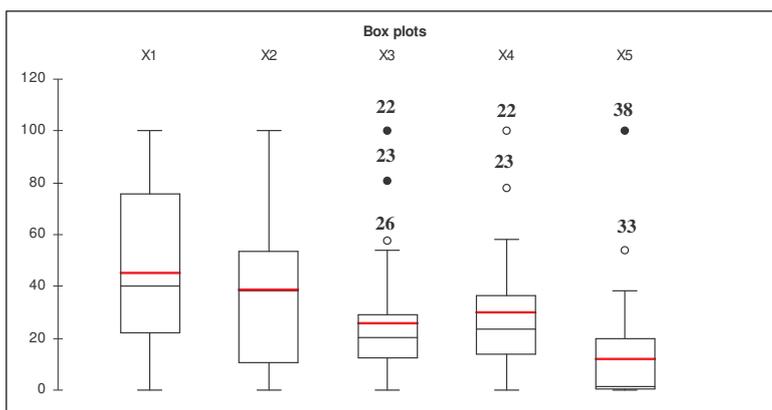


Figure B.3 Boîte à pattes ou à moustaches de la variable X_j .

Ce type de boîtes permet :

- de comparer les distributions des différentes variables à partir de ces cinq paramètres,
- d'illustrer l'étendue de chaque variable à partir de l'intervalle interquartile (plus la boîte est longue, meilleure est la variabilité),
- de détecter les **expériences aberrantes** ou les **erreurs** (de mesure ou de saisie) d'un tableau de données que l'on appelle des **individus atypiques**.



La **Figure B.4** représente un exemple de **boîtes à moustaches** de cinq variables d'une même analyse sur lesquelles a été rajoutée en rouge la moyenne arithmétique. Dans cet exemple, on note la présence de 7 individus atypiques.

Figure B.4 Box Plot de cinq variables centrées réduites

II.2.2. ANALYSE BIDIMENSIONNELLE

L'analyse bidimensionnelle détermine les relations existant entre deux variables. Nous allons nous placer dans le cas où les valeurs prises par ces variables sont quantitatives.

II.2.2.a. La régression simple [8]

De façon schématique, la régression simple est la mise en équation de la relation entre deux variables x et y .

Le problème

On note x_j et y_j , deux variables ; soit x_{ij} et y_{ij} les valeurs prises par ces variables. On cherche à déterminer s'il existe une relation linéaire vérifiée même approximativement par les deux variables, c'est à dire s'il existe deux réels a et b tel que :

$$y_{ij} = ax_{ij} + b + e_{ij} \quad \text{avec } i = 1, \dots, n \text{ et } e_{ij} \text{ un terme résiduel.}$$

La relation entre y_j et x_j sera d'autant plus proche d'une relation linéaire exacte que les valeurs de la série e , c'est à dire les valeurs de e_{ij} seront petites. Algébriquement, on détermine a et b selon le critère des moindres carrés c'est à dire de telle manière que $\sum_{i=1}^n e_{ij}^2$ ait une valeur minimale. Graphiquement (voir

Figure B.5), chaque individu est représenté par un point de coordonnées x_{ij} et y_{ij} dans un repère d'axes x et y . On recherche la droite qui passe « au plus près » du nuage de points. Le terme résiduel e_{ij} est l'écart entre la droite et la valeur observée de y_{ij} .

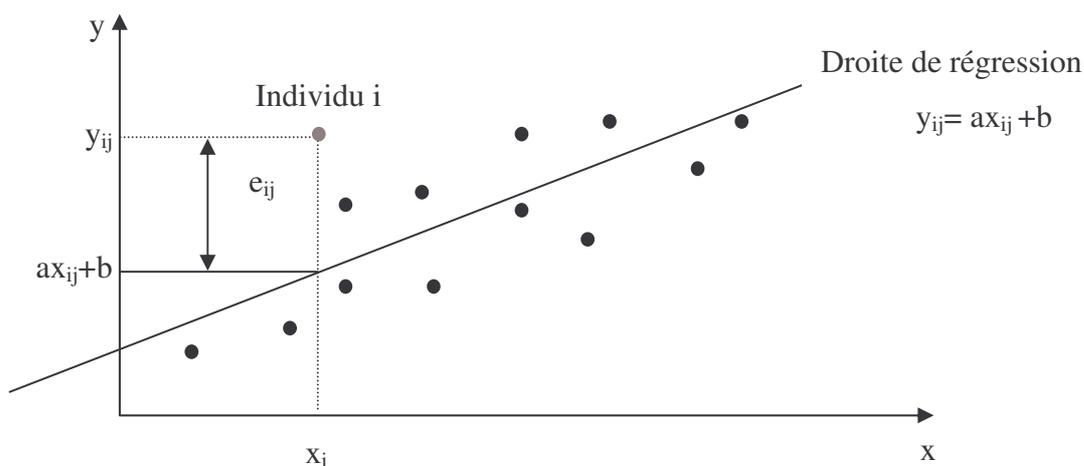


Figure B.5 Droite de régression

Résolution algébrique

La droite de régression des moindres carrés $y_{ij} = ax_{ij} + b + e_{ij}$ est telle que :

$$a = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(y_{ij} - \bar{y}_j)}{n \text{Var}(x)} ; \quad b = \bar{y}_j - a\bar{x}_{ij} ; \quad \sum_{i=1}^n e_{ij} = 0 .$$

Par définition, la covariance entre 2 variables x et y est :

$$\text{Cov}(x, y) = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(y_{ij} - \bar{y}_j)}{n} \quad \text{et donc} \quad a = \frac{\text{Cov}(x, y)}{\text{Var}(x)}$$

Ces définitions nous sont nécessaires pour comprendre au mieux la signification des coefficients de corrélations et de déterminations, qui sont des indicateurs essentiels en analyse de données.

Equation de la variance

L'importance des fluctuations des valeurs prises par la variable y_j est mesurée par la variance notée $\text{Var}(y)$. Les fluctuations de y_j occasionnées par les valeurs prises par la variable x_j sont mesurées par $\text{Var}(ax+b)$ et les fluctuations de y_j ne dépendant pas de x_j sont mesurées par $\text{Var}(e)$. On obtient l'équation suivante, appelée « équation d'analyse de la variance » :

Equation d'analyse de la variance :

$$\text{Var}(y) = \text{Var}(ax+b) + \text{Var}(e)$$

$\text{Var}(y)$: variance totale de y_j

$\text{Var}(ax+b)$: variance expliquée par les variations de x_j

$\text{Var}(e)$: variance des résidus

Grâce à ces définitions, nous allons pouvoir déterminer la formule mathématique des coefficients de corrélation et de détermination.

II.2.2.b. Coefficients de corrélation et de détermination [8]

Formulation

On mesure l'intensité du lien linéaire entre x et y par le rapport de la variance expliquée $\text{Var}(ax+b)$ par la variance totale $\text{Var}(y)$. Ce rapport, appelé coefficient de détermination, est noté $R^2(x, y)$.

La racine carrée de $R^2(x,y)$, affectée par convention du signe de a , donc du signe de $\text{Cov}(x,y)$, est le coefficient de corrélation linéaire entre x et y : $R^2 = \frac{\text{Var}(ax+b)}{\text{Var}(y)} = \frac{a^2 \text{Var}(x)}{\text{Var}(y)}$ d'où $R^2(x,y) = \frac{\text{Cov}(x,y)}{\sigma_x \sigma_y}$.

Le coefficient de détermination prend des valeurs comprises entre 0 et 1, il découle donc que la valeur du coefficient de corrélation est comprise entre -1 et 1 . Notons aussi que ces coefficients sont symétriques par rapport à x et y :

$$R(x, y) = R(y, x).$$

Interprétation géométrique du coefficient de corrélation

Une variable x (variable prenant n valeurs) peut être représentée par un vecteur dans \mathfrak{R}^n (espace des variables). Dans cet espace, le produit scalaire entre deux vecteurs \vec{x} et \vec{y} de coordonnées respectives

$$(x_1, \dots, x_n) \text{ et } (y_1, \dots, y_n) \text{ est : } \vec{x} \cdot \vec{y} = \sum_{i=1}^n x_i y_i .$$

Ce produit scalaire permet de donner une interprétation géométrique simple du coefficient de corrélation : En effet, dans l'espace \mathfrak{R}^n , le cosinus de l'angle entre 2 variables centrés X et Y est égal au coefficient de corrélation entre ces variables (voir **Figure B.6**)

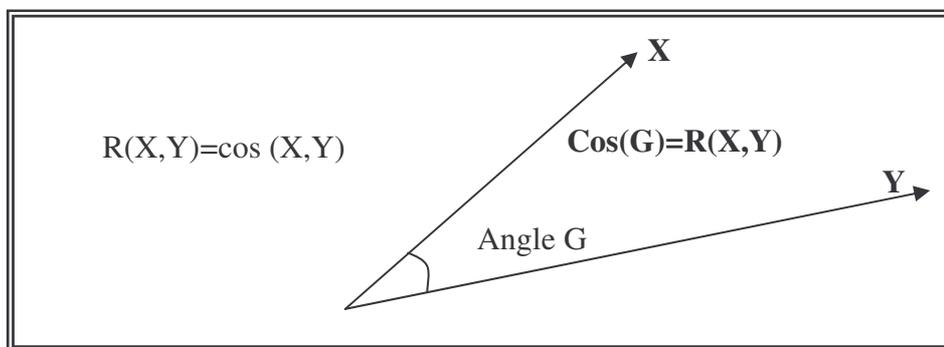


Figure B.6 Coefficient de corrélation dans \mathfrak{R}^n

Interprétation du coefficient de corrélation dans \mathfrak{R}^n

Le coefficient de corrélation $R(X,Y)$ traduit la liaison entre deux variables X et Y (voir **Figure B.7**). Il varie entre -1 et $+1$. La valeur -1 ou $+1$ signifie qu'il existe une relation linéaire exacte, respectivement décroissante ou croissante, entre les deux variables.

Par contre, une valeur faible (proche de 0) pour ce coefficient de corrélation peut traduire, non seulement l'absence de liaison linéaire, mais éventuellement l'existence d'une liaison non linéaire.

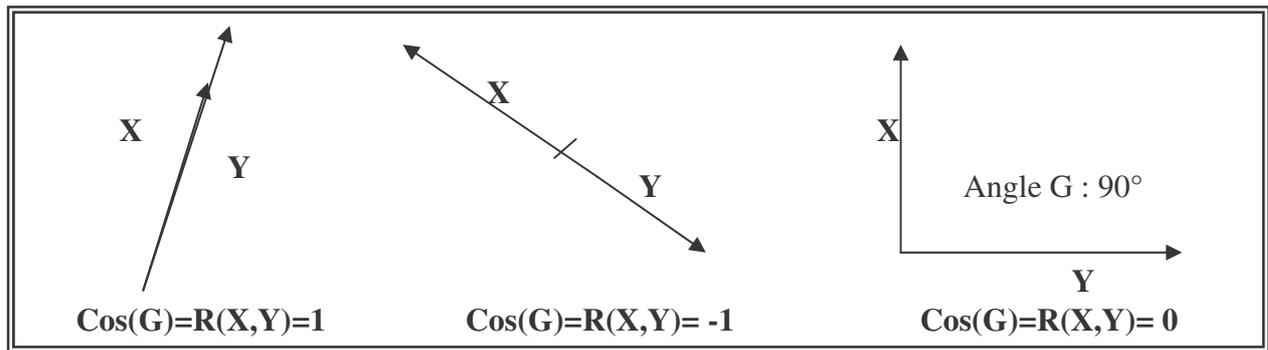


Figure B.7 Interprétation du coefficient de corrélation

Analyse des corrélations ⁽³⁾

Il est possible de faire une première analyse des corrélations grâce à deux méthodes : de façon analytique avec la valeur numérique du coefficient de détermination (que l'on appelle la force de liaison) ou de façon graphique avec l'observation de l'allure du nuage de points en fonction des valeurs du coefficient de corrélation (matrice X Y).

□ LA FORCE DE LIAISON

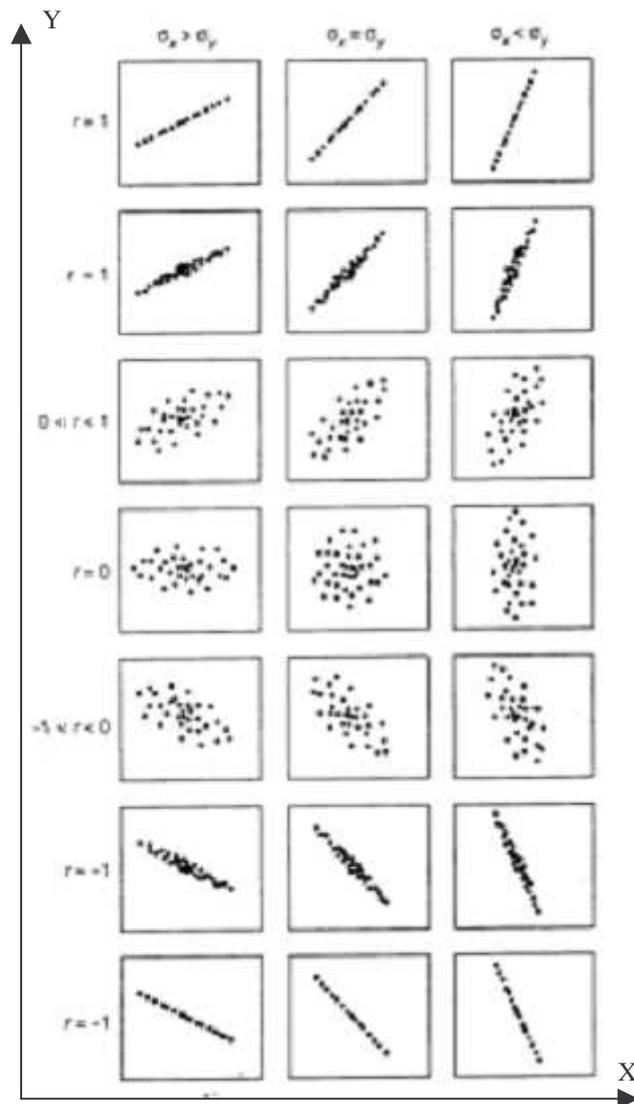
La valeur absolue du coefficient de corrélation permet de mesurer le « risque d'erreur » en affirmant qu'il existe une liaison statistiquement significative entre les deux variables concernées. Mais, il ne permet pas d'évaluer directement la force de cette liaison. Si on appelle X la variable causale (celle qui détermine l'autre) et Y la variable à expliquer, il est rarissime comme le montre le **Tableau B.4** que l'on ait un coefficient de corrélation de 1 (ou -1) excepté les valeurs diagonales. Si c'était le cas, cela voudrait dire que toutes les variations de Y seraient expliquées par celles de X. Pour connaître l'importance de la variable X sur Y, il faut calculer le **coefficient de détermination** : R^2 . Cet indice exprime directement la part (en pourcentage) de la variance de Y expliquée par celle de X.

Tableau B.4 Matrice des corrélations

	X1	X2	X3	X4	X5
X1	1	-0,084	-0,468	-0,229	0,690
X2	-0,084	1	-0,183	-0,388	-0,532
X3	-0,468	-0,183	1	0,770	-0,193
X4	-0,229	-0,388	0,770	1	0,028
X5	0,690	-0,532	-0,193	0,028	1

- $R(X3, X1) = -0,468$; si on élève cette valeur au carré on obtient : $R^2(X3, X1) = 0,218$ ce qui signifie que 21,8 % des variations de X3 sont expliquées par les variations de la variable X1. Ces deux variables sont **faiblement corrélées**.

- $R(X5, X1) = 0,69$; $R^2(X5, X1) = 0,476$ ce qui signifie que 47,6 % des variations de X5 sont expliquées par les variations de X1. Dans ce cas, on peut dire que X5 et X1 sont **fortement corrélés**.
- $R(X5, X4) = 0,028$; $R^2(X5, X4) = 0,00078$ (proche de 0) ce qui signifie que les variations de la variables X5 ne dépendent pas des variations de la variables X4. En d'autres termes, X5 et X4 sont des variables **indépendantes**.



□ L'OBSERVATION DU NUAGE DE POINTS

L'observation du nuage de points de deux variables X et Y nous donne également une idée sur la corrélation de ces variables. La **Figure B.8** résume les différentes configurations existantes en fonction du coefficient de corrélation et des écarts types de X (σ_X) et de Y (σ_Y).

Rappels :

$$-1 \leq R \leq 1$$

X et Y **indépendants** : $Cov(X, Y) = 0$ donc

R = 0 (absence de corrélation linéaire)

$|R| \geq 0,9$: forte corrélation linéaire

R=1 ou **R=-1** : nuage de points équivalent à une droite

R < 0 : Y décroît quand X croît

(la qualité de l'ajustement est mesurée par le coefficient de détermination R^2)

Figure B.8 Allures du nuage de points en fonction du coefficient de corrélation

Un outil d'analyse de données, appelé la **matrice XY** ou encore le « **Scatters Plot** » permet de regarder rapidement à titre de **pré-analyse** ces nuages de points sur un même graphique.

⁽³⁾ Francis MARCHAND, *Initiation aux statistiques - Notes de cours, Université de Limoges, 2000*

La matrice XY est un complément d'une analyse des corrélations plus poussée tel que l'analyse en composante principale (chapitre III). A titre d'exemple, la **Figure B.9** représente la matrice XY correspondant aux données du **Tableau B.2**.

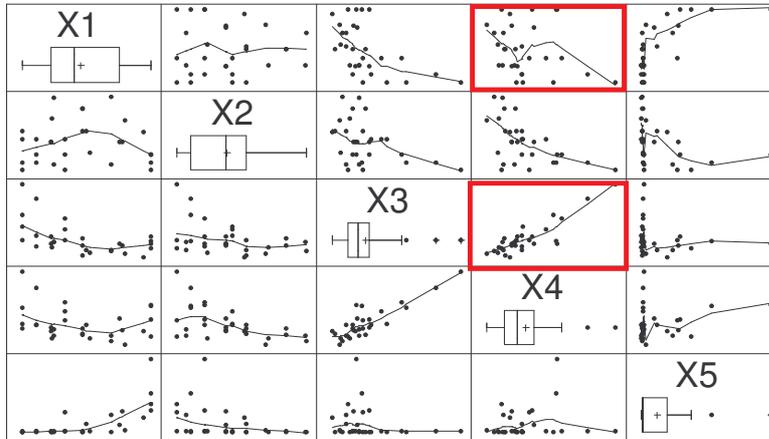


Figure B.9
Matrice XY

II.3. ORGANIGRAMME

Pour bien saisir les démarches et l'enchaînement des différentes étapes dans l'analyse de données et plus généralement pour une problématique, on résume les différentes étapes sous forme d'organigramme (**Figure B.10**):

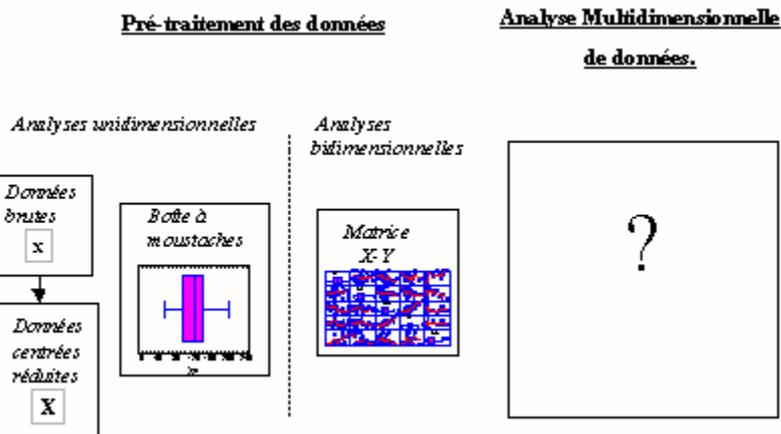


Figure B.10 Différentes étapes de l'analyse de données

Après avoir réalisé le pré-traitement des données, l'étape suivante que nous allons expliquer dans les chapitres suivants, concerne l'analyse multidimensionnelle de données.

Chapitre III L'analyse en composante principale

(A.C.P)

III.1. INTRODUCTION

L'Analyse en Composante Principale (A.C.P) est une méthode d'analyse de données multidimensionnelles et plus particulièrement **une analyse factorielle**. Elle est née d'un besoin qui s'impose comme une évidence dans le domaine scientifique où, aujourd'hui on est appelé à manipuler de grands ensembles de données. Le propre de l'A.C.P. est de raisonner sur un nombre quelconque de variables : dans toute problématique ayant des résultats où interviennent de très nombreuses variables, il est intéressant de pouvoir les traiter ensemble et non plus seulement deux à deux [10]. Ce type d'analyse représente, sans doute, la meilleure introduction aux méthodes offertes par la statistique exploratoire des données multidimensionnelles. Il a fallu attendre la démocratisation des outils informatiques (à partir des années 60) et la diffusion de logiciels spécialisés pour rendre accessibles, à un large public, différentes méthodologies comme l'A.C.P. permettant de résumer, souvent sous forme de représentations graphiques utilisant des projections, de grands tableaux de données.

III.2. PRESENTATION GENERALE SUR L'A.C.P.

III.2.1. INTRODUCTION

Dans la plupart des applications, on observe non pas une variable par individu, mais un nombre p souvent élevé. L'étude séparée de chacune de ces variables selon les techniques exposées précédemment est une phase indispensable dans le processus de dépouillement des données mais tout à fait insuffisante : en effet l'étude séparée de chaque variable laisse de côté les liaisons qui peuvent exister entre elles et qui sont souvent l'aspect le plus important [11]. L'A.C.P. a donc pour objet de résumer de grands ensembles de données quantitatives de façon claire et visuelle.

III.2.2. QUELS TYPES DE PROBLEMES PEUVENT ETRE ABORDES PAR L'A.C.P. ?

Les applications de l'A.C.P. sont très nombreuses. Cette méthode permet de rechercher les principales variables explicatives d'une expérimentation mais également d'identifier les corrélations existantes entre ces différentes variables. Elle permet également la recherche des familles de variables et d'individus. Ainsi, il sera possible de faire des groupes d'individus aux performances voisines.

L'A.C.P. va nous permettre d'obtenir des graphiques où l'on pourra observer, aussi objectivement que possible, les ressemblances et dissemblances de ces individus. Elle peut s'avérer être une première étape à l'emploi d'autres méthodes statistiques plus adaptées à notre problème [8][12].

On utilise l'A.C.P. pour répondre par exemple à ces questions :

Existe-il des individus « atypiques » dans le tableau de données ? Et/ou comment se répartissent les individus ? Quels sont ceux qui se ressemblent ? Quels sont ceux qui sont dissemblables ? Comment peut-on regrouper des individus qui ont des comportements homogènes ?

Comment se structurent les variables ? Quelles sont celles qui sont associées ? Quelles sont celles qui ne le sont pas ? Quelles sont celles qui vont dans le même sens ? Quelles sont celles qui s'opposent ? Quels tableaux de données peut-on traiter avec l'A.C.P. ?

III.2.3. QUELS TABLEAUX DE DONNEES PEUT-ON TRAITER AVEC L'A.C.P. ?

L'A.C.P. ne s'applique qu'à un tableau de variables quantitatives. Voici quelques exemples de tableaux pouvant être analysés par cette méthode :

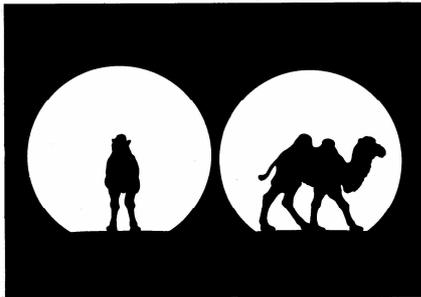
- les tableaux de mesures : pour ce type de tableau, les valeurs des variables sont obtenues à partir d'appareils de mesure,
- les tableaux de notes : pour ce type de tableau, les variables sont obtenus à partir de notations. Ces notes peuvent être, par exemple, une caractérisation esthétique que l'on évalue entre 1 et 10 (1 = pas beau ; 10 = très beau). Ce sont des « **variables qualitatives ordinales** » qui peuvent être assimilées à des variables quantitatives à condition que leur intervalle de variation soit suffisamment grand (8 à 10 classes de notes au moins). Il faut également admettre que l'œil remplace un instrument de mesure.

III.2.4. PRINCIPE DE L'A.C.P. [13]

Les données sont rangées dans un tableau comportant souvent un grand nombre d'individus n , et un grand nombre de variables p ; la simple lecture de ce tableau ne permet pas de saisir l'essentiel des informations qu'il contient. L'analyse en composante principale est une méthode de projection qui résume et décrit ce tableau de données. Si un tableau de données comporte p variables quantitatives, on dira que les individus sont représentés dans un « **espace à p dimensions** ».

Il est inutile de préciser qu'il est impossible de « voir » dans un tel espace. Le but de l'A.C.P. est de trouver des « espaces de dimensions plus petites » dans lesquels il est possible d'observer « au mieux » les individus.

Pour des raisons évidentes de facilité de visualisation, les espaces retenus seront à 1 dimension ou le plus souvent à 2 dimensions. Ce processus est utilisé inconsciemment lorsqu'on réalise des photos : on passe bien d'un espace à 3 dimensions à un espace à 2 dimensions : la photo.



Cependant, comme on peut le voir sur la **Figure B.11**, selon l'angle sous lequel est pris un sujet, les photographies n'apportent pas la même information.

Figure B.11 Dessin emprunté à **J.P Fénelon**

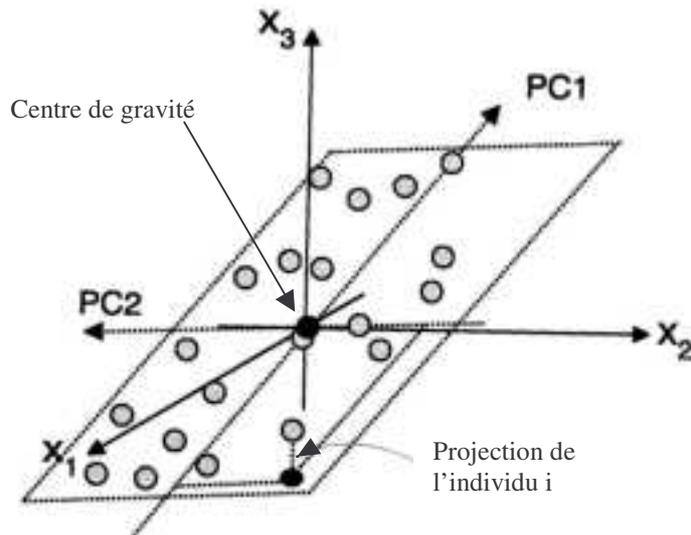
III.3. DEMARCHE D'UNE ANALYSE FACTORIELLE

III.3.1. CONSTRUCTION DES AXES PRINCIPAUX, DEMARCHE GEOMETRIQUE

Les droites et les plans issus d'une A.C.P. ne sont pas réalisés avec les variables initiales mais avec des « indices synthétiques » obtenus à partir de « combinaisons linéaires » des variables initiales. Parmi tous les indices possibles, l'A.C.P. recherche d'abord celui qui permet de voir au mieux les individus c'est-à-dire celui pour lequel la variance des individus est maximale : cet indice est appelé **1^{ère} composante principale** portée par le 1^{er} axe principal appelé axe 1 ou **PC1 (Principal Component 1)**. Une certaine proportion de la variation totale des individus est expliquée par cette composante principale. Ensuite une **deuxième composante principale** est recherchée, et ceci sous trois conditions :

- **$R(\text{PC1}, \text{PC2}) = 0$** (*on rappelle que R est le coefficient de corrélation*) ; En effet, le but de l'A.C.P. étant de former un nouveau repère orthonormé, la condition nécessaire est donc la perpendicularité des axes 1 et 2.
- **la variance (PC2) doit être, à son tour, la plus grande possible.** Ainsi, cette deuxième composante principale fournit la plus grande information possible complémentaire à la première. Le processus se déroule jusqu'à l'obtention de la p^{ième} et dernière composante principale, les parts d'informations expliquées par chacune d'elles devenant de plus en plus faibles (avec p variables, on obtient p composantes principales),
- L'intersection de tous les axes principaux se situe au centre de gravité du nuage de points.

En conclusion, la phase essentielle de l'A.C.P. consiste à transformer ces p variables quantitatives initiales, toutes plus ou moins corrélées entre elles, en p nouvelles variables quantitatives, non corrélées, appelées composantes principales.



La **Figure B.12** illustre la projection des individus dans le plan principal formé par les deux premières composantes principales PC1 et PC2. Soulignons que dans cet exemple nous avons une projection en trois dimensions (repère X_1, X_2, X_3) au lieu de p dimensions, impossible à représenter.

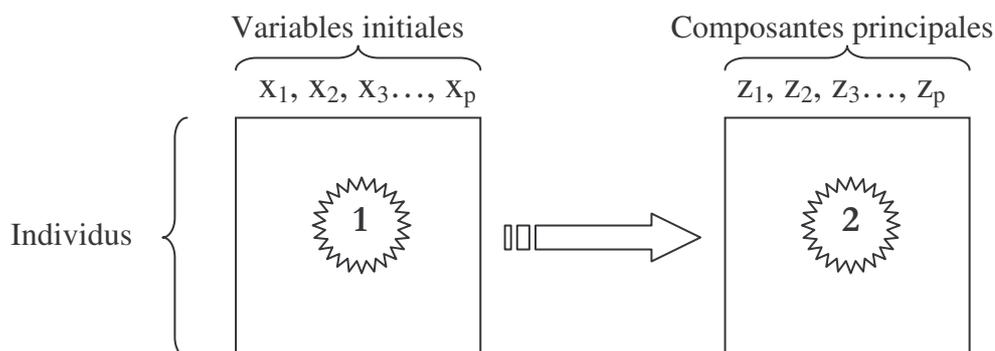
Figure B.12 Projection des individus dans le plan principal

III.3.2. NOTION D'INERTIE : DEMARCHE ANALYTIQUE

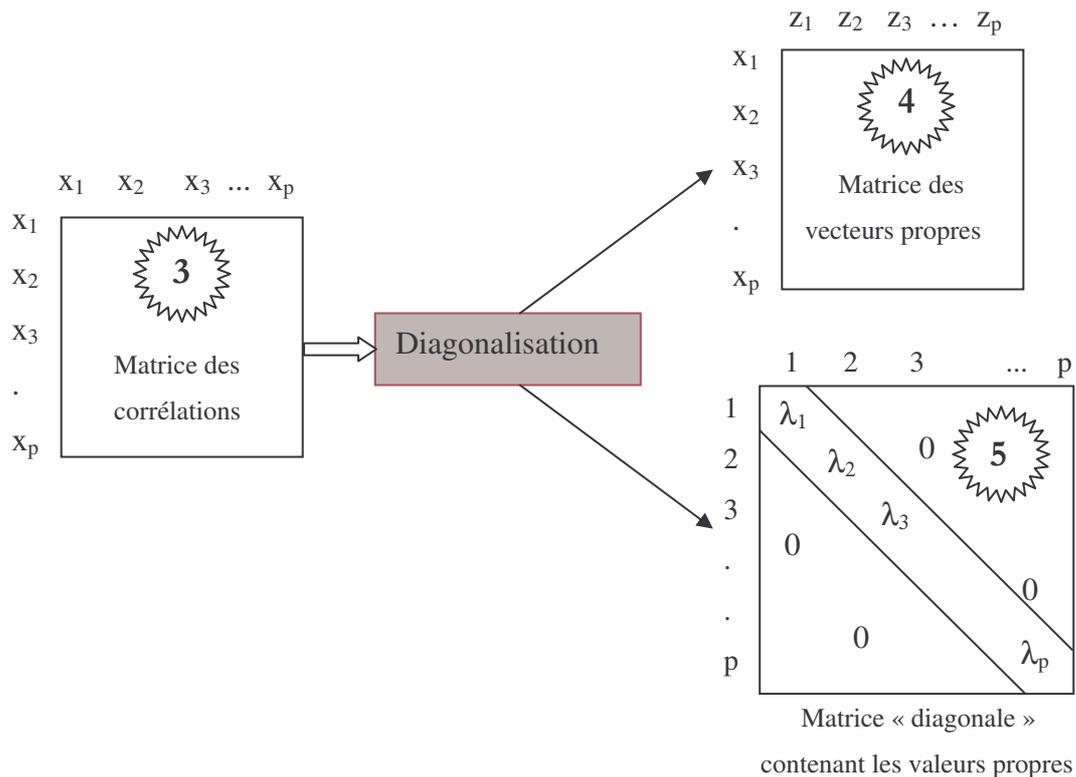
III.3.2.a. Calcul matriciel

Dans l'analyse de données, d'un point de vue un peu moins géométrique, on opte en faveur de la notion d'**inertie** d'un nuage de point représentant les individus. Pour avoir en détail le principe et la démonstration mathématique il est préférable de se référer à la bibliographie [12].

Nous allons essayer d'expliquer cette notion d'inertie, parfois considérée comme complexe, de façon schématique. Dans le paragraphe précédent, nous avons vu que l'A.C.P. consistait à transformer les p variables quantitatives initiales *inter corrélées* (1) en nouvelles variables *non corrélées* (2) appelées **composantes principales** comme l'indique le schéma ci dessous.



Pour obtenir ce résultat de façon analytique (et non plus de façon géométrique) nous allons procéder en deux étapes : élaborer la **matrice des corrélations (3)** et faire une opération mathématique qui s'appelle la **diagonalisation [14]**. Cette opération permet d'obtenir une **matrice de vecteurs propres (4)**, qui contient les coefficients des combinaisons linéaires des p variables initiales actives et la **matrice « diagonale » de valeurs propres (5)**, représentant les variances des individus sur les axes principaux correspondants comme on peut le voir ci dessous :



III.3.2.b. Aide à l'interprétation

Nous pouvons tirer de l'information des matrices 4 et 5. En effet :

- les **valeurs propres** représentent les **variances** des individus sur les axes correspondants. On rappelle que la **variance** des individus, lorsque les **variables** sont **centrées réduites**, est égale à **1** (voir **Tableau B.3**). Dans ce cas-là, si l'on additionne les valeurs numériques des variances, on obtient le nombre de variables p (par exemple, pour les données du **Tableau B.3**, p étant égal à 5, la somme des variances = 5). C'est grâce à ces définitions et à la notion d'inertie que l'on calcule la **part d'information** initiale, « visible » sur chaque axe, en faisant le rapport de chaque valeur propre (λ_i) à la somme de toutes les variances initiales (p).
- les **vecteurs propres** permettent le calcul des composantes principales et donc le calcul des coordonnées des variables et des individus sur les nouveaux axes principaux (voir paragraphe suivant).

III.4. PRESENTATION DE L'A.C.P. A PARTIR D'UN EXEMPLE

Dans cette partie, nous allons expliquer ce qu'il est nécessaire de comprendre pour effectuer une analyse en composante principale. Nous avons choisi de réaliser une A.C.P à partir des données du **Tableau B.2**.

III.4.1. BILAN DES ANALYSES PRECEDENTES

Dans le chapitre II, nous avons réalisé une pré-étude des données reportées dans le **Tableau B.2**. Nous avons déterminé à l'aide des boîtes à moustaches, les individus atypiques. Il est conseillé de déclarer ces individus en tant qu'**individu supplémentaire dans l'A.C.P.**. Ainsi, ils figureront de manière descriptive dans l'analyse sans participer au calcul des axes principaux et donc ne déstabiliseront pas l'analyse. Le **Tableau B.5** regroupe les 5 individus supplémentaires décelés dans le pré-traitement des données (voir **Figure B.4**).

Tableau B.5 Individus supplémentaires

Nombre	X1	X2	X3	X4	X5
22	1,0	0	85,0	370,0	22
23	2,0	50	70,0	300,0	13
26	1,0	50	52,3	216,0	10
33	10,0	50	25,3	224,0	594
38	10,0	100	21,8	225,0	1100

A présent, nous pouvons commencer à chercher les plans (donc les axes principaux) dans lesquels nous allons visualiser les variables et les individus.

III.4.2. PART D'INFORMATION PORTEE PAR LES AXES

Dans notre exemple, nous avons 5 variables initiales ($p = 5$). On calcule alors les pourcentages de variations expliqués (part d'information) par les axes principaux avec les valeurs diagonales contenues dans le **Tableau B.6** (voir l'explication dans le paragraphe III.3.2.b).

Tableau B.6 Matrice diagonale contenant les valeurs propres

	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
Axe 1	2,197	0	0	0	0
Axe 2	0	1,862	0	0	0
Axe 3	0	0	0,585	0	0
Axe 4	0	0	0	0,219	0
Axe 5	0	0	0	0	0,136

On appelle **Z** les nouvelles variables produites par l'A.C.P. (donc les nouveaux axes) : l'axe 1 correspond alors à Z1, l'axe 2 à Z2, etc.... Le pourcentage d'information porté par chaque axe (**Tableau B.7**) est calculé ci-dessous. Pour l'axe 1 (Z1), on fait le calcul suivant :

$$100 \times \frac{\lambda_1}{p} = 100 \times \frac{2,197}{5} = 43,945 \%$$

$$\text{et pour l'axe 2 (Z2) : } 100 \times \frac{1,862}{5} = 37,246 \%$$

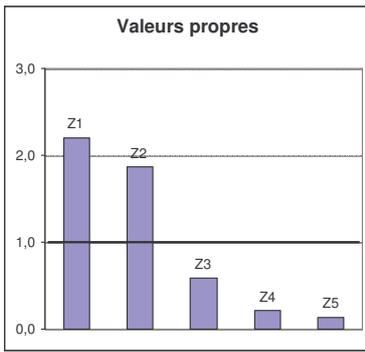
Si on cumule les informations apportées par les axes 1 et 2, on obtient la part d'information apportée par le plan 1/2. Les différentes composantes principales étant non corrélées, la part d'information expliquée par deux axes donnés (espace à **deux dimensions**) est égale à la simple addition des parts expliquées par chacun de ces axes. Le **pourcentage d'information cumulé**, pour les axes 1 et 2, est donc $43,945 + 37,246 = 81,191 \%$ (voir **Tableau B.7**).

Tableau B.7 Pourcentage d'information (variance, variance cumulé)

	Z1	Z2	Z3	Z4	Z5
Valeur propre	2,197	1,862	0,585	0,219	0,136
% variance	43,945	37,246	11,709	4,384	2,716
% cumul	43,945	81,191	92,900	97,284	100,000

Remarques :

- ⊙ l'axe (Z1) explique à lui seul presque la moitié de l'information initiale (44%),
- ⊙ 2 axes (Z1 et Z2) expliquent les 4/5 de l'information (81 %),
- ⊙ 3 axes expliquent plus de 9/10 (93 %),
- ⊙ et 4 axes n'expliquent pas beaucoup plus (97 %).



De plus, si la part de variation expliquée par un axe (valeur propre) est supérieure ou égale à celle de la variable initiale (égale à 1 (voir tableau 4)), on choisira de regarder les projections des individus et des variables sur cet axe (voir **Figure B.13**). Donc ici, les axes Z1 et Z2 définiront le plan de représentation appelé plan principal. Ces règles doivent s'adapter à chaque cas particulier et sont flexibles.

Figure B.13 Histogramme des valeurs propres

III.4.3. NOUVELLES COORDONNEES DES INDIVIDUS ET DES VARIABLES

III.4.3.a. Nouvelles coordonnées des individus

On cherche à présent à calculer les coordonnées des individus sur les nouveaux axes (Z1, Z2, Z3, Z4 et Z5). Comme nous l'avons expliqué dans le paragraphe III.3.2, les vecteurs propres obtenus précédemment permettent de calculer les nouvelles coordonnées des variables dans le repère des composantes principales. La **Figure B.14** montre que ces nouvelles coordonnées [Z] sont obtenues en multipliant la matrice des vecteurs propres [V] avec la matrice des coordonnées centrées réduites [X] (**Tableau B.3**).

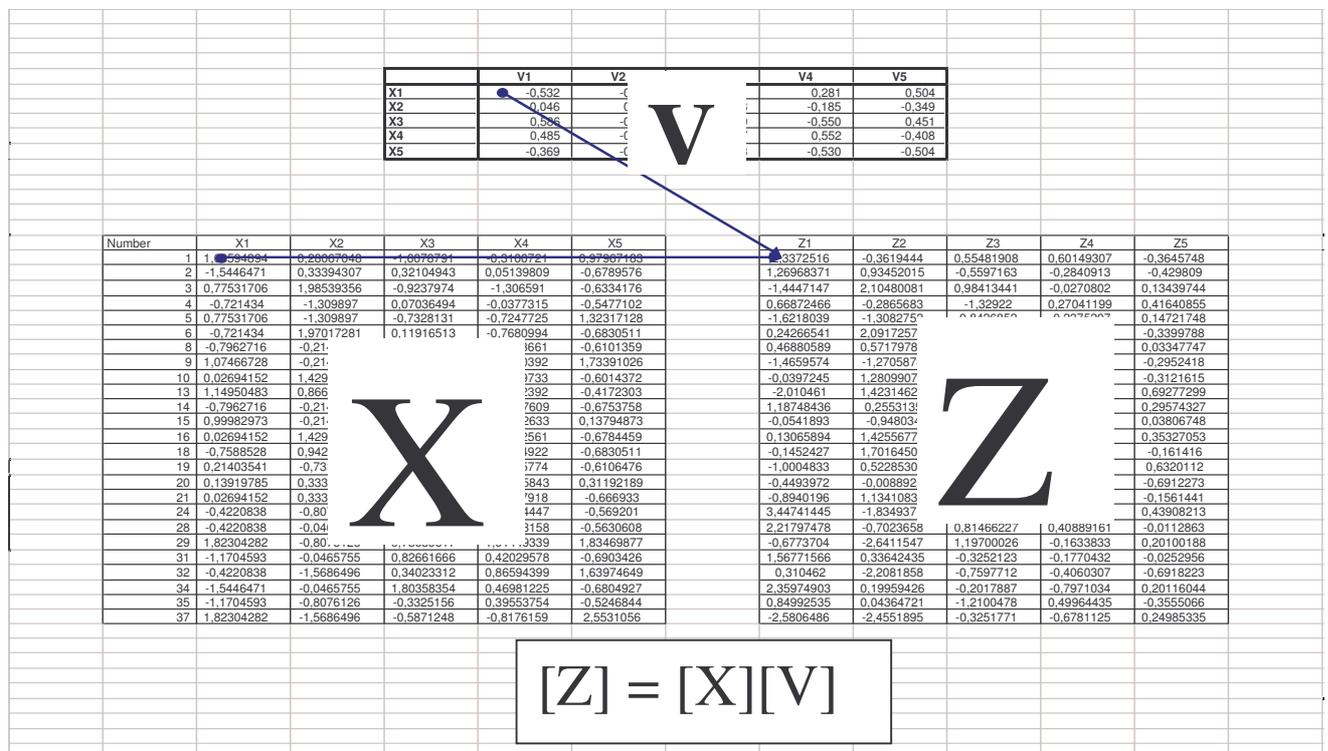


Figure B.14 Calcul dans Excel des nouvelles coordonnées des individus.

Avec :

V : matrice des vecteurs propres,

X : matrice des coordonnées centrées réduites des individus,

Z : matrice des nouvelles coordonnées des individus sur les nouveaux axes.

III.4.3.b. Nouvelles coordonnées des variables

Pour calculer les nouvelles coordonnées des variables, il faut multiplier la matrice des vecteurs propres par la racine carrée des valeurs propres (voir **Figure B.15**).

Les démonstrations et les formules mathématiques associées à ces calculs, sont entièrement décrites dans les références bibliographiques [8] et [14].

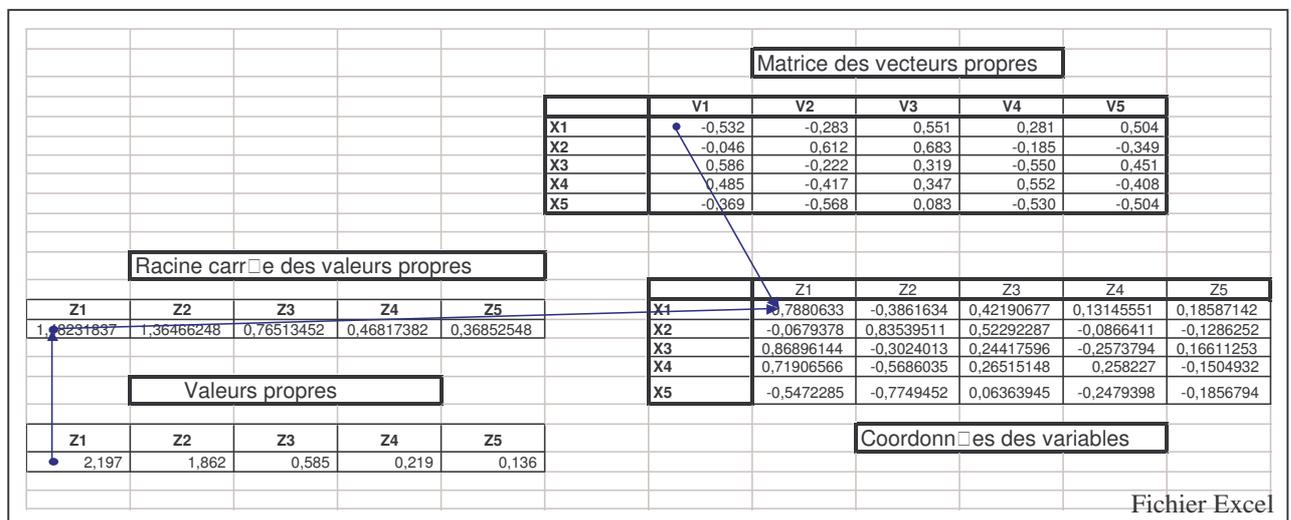


Figure B.15 Calcul dans Excel des nouvelles coordonnées des variables

Grâce à cette matrice des coordonnées des variables, on construira par la suite, le cercle des corrélations.

III.4.4. CONTRIBUTION DES INDIVIDUS AU CALCUL DES AXES

La contribution (CTR) des individus i à la construction d'un axe j , se calcule avec la formule ci-dessous :

$$CTR(i, \text{axe } j) = \frac{(Z_{ij})^2}{N \times \lambda_j}$$

Ce sont donc les individus qui ont les plus fortes coordonnées, en valeur absolue, qui contribuent le plus à la formation des axes.

Deux cas sont alors possibles :

- soit un axe est essentiellement dû à quelques individus que l'on peut qualifier de suspects. Il faut alors recommencer l'analyse après avoir pris soin de les retirer du calcul des axes en les déclarant, dans l'analyse, en individus supplémentaires,
- soit un axe est dû à un ensemble assez homogène d'individus ; on pourra alors le qualifier de « stable » et lui accorder une certaine signification
- Nous pouvons représenter sur un histogramme (**Figure B.16**), les valeurs absolues des coordonnées Z_{ij} des individus (dans le repère de l'A.C.P), afin de voir en globalité les contributions des individus pour chaque axe. Par exemple, on peut voir que l'individu 24 contribue fortement à la formation de l'axe 1. Il sera donc conseillé de le retirer de l'analyse en le déclarant en individu supplémentaire.

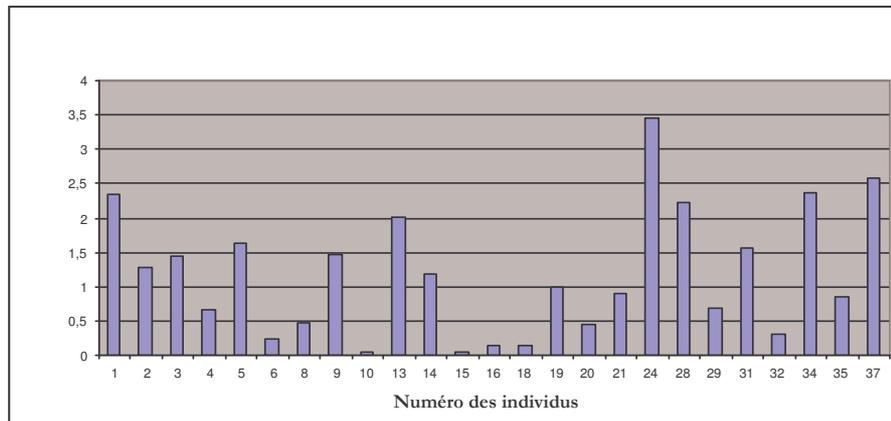


Figure B.16 Contribution des individus à l'axe 1

III.4.5. QUALITE DE REPRESENTATION DES INDIVIDUS

Par la suite, nous allons nous intéresser aux proximités et aux éloignements des individus entre eux. Mais, n'oublions pas que les plans principaux de l'A.C.P. ne permettent pas d'avoir une vue parfaite de la réalité et qu'il faut se méfier des déformations car le pourcentage d'information cumulé n'est pas égal à 100 %. Si un nuage de points est globalement bien représenté dans un plan, inévitablement certains points y sont mal représentés.

Pour savoir comment les individus sont représentés sur les axes (ou dans le plan principal formé par les axes PC1 et PC2), on regarde si chacun des points projetés est proche de celui-ci ou non.

La **Figure B.17** représente un exemple de projection des individus sur l'axe principal 1 (PC1). Dans le **Tableau B.8** la valeur de l'angle α (et plus particulièrement la valeur du cosinus carré de cet angle) nous donne l'information sur la qualité de représentation de cet individu.

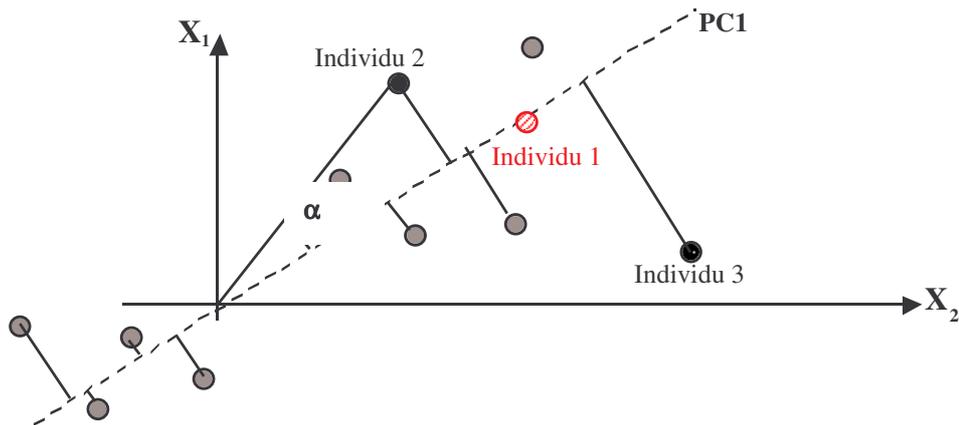


Figure B.17 Projection des individus sur l'axe 1 (PC1)

Dans l'exemple de la **Figure B.17**, si on appelle respectivement α_1 , α_2 et α_3 les angles entre PC1 et les individus 1, 2 et 3 ($\alpha_1 = 0^\circ$; $\alpha_2 = 20^\circ$; $\alpha_3 = 30^\circ$), les valeurs des cosinus carrés correspondants ($\cos^2(\alpha_1) = 1$; $\cos^2(\alpha_2) = 0,88$; $\cos^2(\alpha_3) = 0,86$) montrent que les individus sont bien représentés sur l'axe 1 (PC1).

De manière générale pour connaître la qualité de représentation d'un individu, on postule les règles énoncées dans le **Tableau B.8** ci dessous.

Tableau B.8 Représentation des individus sur l'axe 1

Valeur de $(\cos \alpha)^2$	Signification
Autour de 1	Très bonne représentation
Supérieur à 0,5	Représentation acceptable
Inférieur à 0,5	Mauvaise représentation

De plus, il est intéressant d'additionner les cosinus carrés de chaque axe pour visualiser la qualité de représentation des individus dans un plan. Ainsi, on obtient pour le plan principal (PC1-PC2) le graphique de la **Figure B.18**; chaque expérience est représentée par un point de surface proportionnelle à la somme des cosinus carrés de chacun des axes 1 et 2. Donc, plus le point est gros, plus on peut conclure que le point est bien représenté.

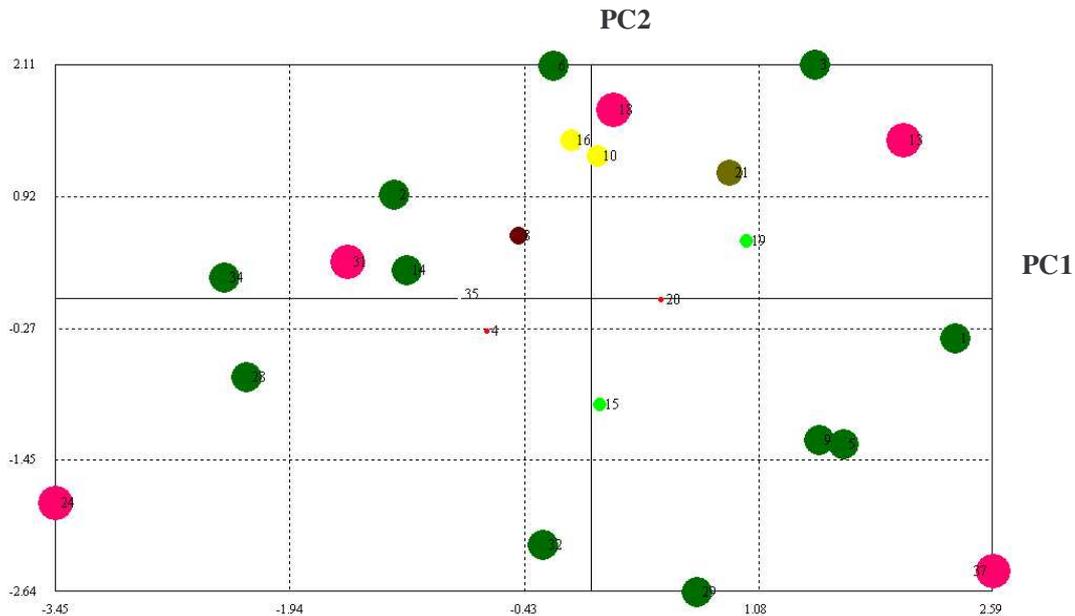


Figure B.18 Représentation des individus dans le plan principal (PC1-PC2)

III.4.6. OUTILS GRAPHIQUES : LE CERCLE DES CORRELATIONS ET LE BIPLLOT

III.4.6.a. Le cercle des corrélations

Définition

Le cercle des corrélations (voir **Figure B.19**) qui représente les **corrélations** entre les variables initiales (x_1, x_2, \dots, x_p) est la représentation la plus utilisée. Il indique l'organisation des variables dans l'espace des composantes principales choisi (plan principal) et permet de répondre à la question suivante : « *Comment les variables sont-elles structurées ?* ».

Pour le construire, on trace un cercle de rayon 1 à l'intérieur duquel on représente les variables (dont les coordonnées sont reportées dans le **Tableau B.9**) dans le plan principal ($Z1 - Z2$) appelé également (PC1 - PC2). On rappelle que les coordonnées des variables représentent aussi les corrélations entre les variables et les axes. Par exemple, les coefficients de corrélation entre la variable $X1$ et les axes $Z1$ et $Z2$ sont respectivement égaux à : $R(X1, Z1) = -0,788$; $R(X1, Z2) = -0,386$.

Dans le cas où, ni le cercle, ni les vecteurs ne sont tracés, il est fortement recommandé de compléter la représentation en reliant le centre du cercle des corrélations aux différentes coordonnées des variables. Ces vecteurs nous permettront par la suite d'interpréter le cercle des corrélations.

Tableau B.9 Coordonnées des variables sur les axes Z1-Z2

	X1	X2	X3	X4	X5
Z1 (Anglais PC1)	-0,7880633	-0,0679378	0,86896144	0,71906566	-0,5472285
Z2 (Anglais PC2)	-0,3861634	0,83539511	-0,3024013	-0,5686035	-0,7749452

Qualité de représentation des variables

Avant d'analyser le cercle des corrélations, il est important de vérifier si les variables sont bien représentées sur le graphique. La qualité de représentation d'une variable sur un axe se mesure en élevant au carré le coefficient de corrélation entre l'axe et la variable (voir **Tableau B.10**). Pour obtenir la part d'information d'une variable expliquée par k axes, on fait la somme des carrés des coefficients de corrélation des k axes. Par exemple, les axes Z1 et Z2 expliquent 77 % des variations de la variable X1.

Tableau B.10 Qualité de représentation des variables dans le plan Z1-Z2

	Z1	Coef. De corrélation au carré	Z2	Coef. De corrélation au carré	Part d'information
X1	-0,788	0,621	-0,386	0,149	77%
X2	-0,068	0,005	0,835	0,698	70,30%
X3	0,869	0,755	-0,302	0,091	84,70%
X4	0,719	0,517	-0,569	0,323	84%
X5	-0,547	0,299	-0,775	0,601	90%

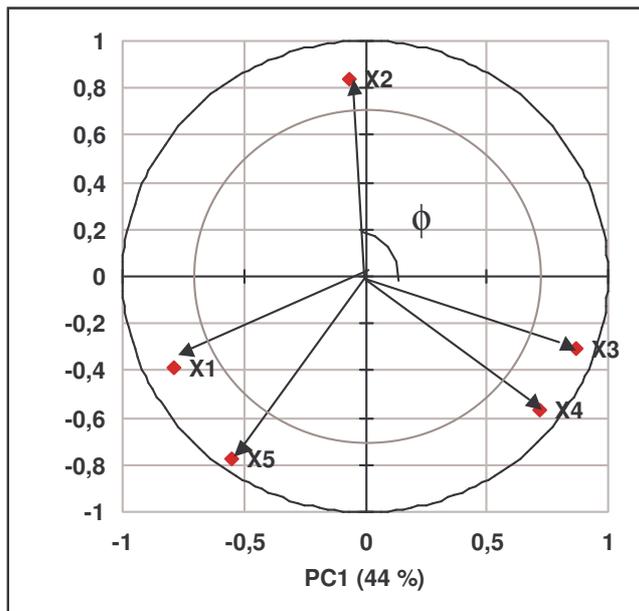
Il est souvent convenu de dire que :

- si la part d'information > 70 % : la variable est très bien représentée.
- si la part d'information est inférieure à 70 %, les variables peuvent être, **moyennement bien** représentées (par exemple 60 %) ou **médiocrement** représentées (par exemple 30%). Dans ce dernier cas, on ne pourra rien conclure sur les relations entre cette variable et les autres variables. Il sera donc nécessaire de visualiser cette variable dans un autre plan où elle sera mieux représentée.

Afin de voir directement sur le cercle des corrélations si les variables sont bien ou médiocrement représentées, on trace **un cercle de rayon 0,7** (voir **Figure B.19**). Ainsi, on n'analysera que les variables se situant à l'extérieur de ce cercle. Dans l'exemple du **Tableau B.10**, on constate que toutes les variables sont bien représentées puisque leur part d'information dans le plan principal choisi (PC1, PC2) est supérieure à 70 %. On pourra alors interpréter toutes les corrélations entre ces variables.

Interprétation

Dans notre exemple, les variables les plus liées à l'axe 1 (Z1) sont X1 et X3. Les vecteurs correspondant à ces deux variables auront la même direction mais seront de sens opposé car les coefficients de corrélation n'ont pas le même signe. Cela signifie que lorsque l'une augmente, l'autre diminue. Les variables X5 et X2 sont les plus liées à l'axe 2 (Z2), leur sens de variation s'oppose également. En règle générale, la relation existant entre les différentes variables est définie par la valeur de l'angle ϕ :



- X1 et X5 sont positivement corrélées.
- X1 et X3 sont inversement corrélées.
- X2 est inversement corrélée à X5 et X4.
- X5 et X4 sont indépendantes.
- X3 et X4 sont positivement corrélées.

Figure B.19 Cercle des corrélations dans le plan principal (PC1-PC2)

- $0^\circ < \text{valeur absolue de } \phi < 90^\circ$: les variables sont **corrélées dans le même sens**. Par exemple, quand la valeur de X1 augmente, la valeur de X5 augmente aussi.

- $90^\circ < \text{valeur absolue de } \phi < 180^\circ$: les variables sont **inversement** corrélées. Quand X1 augmente, X3 diminue. De la même manière quand X2 augmente X5 diminue.

- ϕ proche de 90° : les variables sont **indépendantes**, c'est-à-dire non corrélées. Par exemple, X5 et X4 varient de manière indépendante, il n'existe aucune relation entre ces deux variables.

III.4.6.b. Le Biplot

Afin de connaître quelles sont les variables qui ont joué un rôle prépondérant dans la constitution des composantes principales et donc celles qui expliquent la représentation des individus, on reporte simultanément les individus et les variables dans un même plan appelé Biplot (**Figure B.20**).

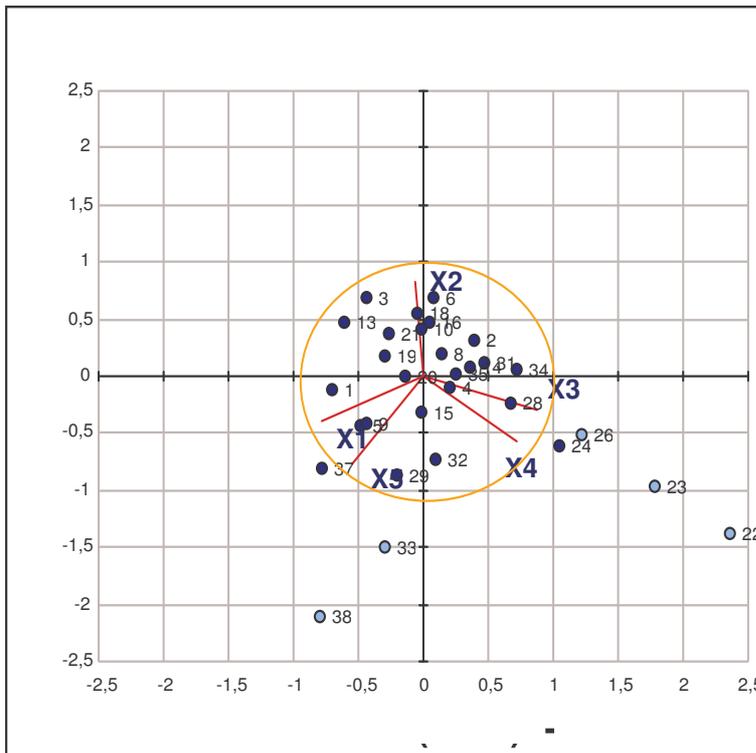


Figure B.20 Biplot

De façon à extraire le maximum d'informations, on étire ce graphique comme on peut le voir sur la **Figure B.21**. Ce graphique apporte des informations sur la répartition des individus, sur les familles d'individus et sur les domaines qui n'ont pas encore été explorés.

Tout d'abord, nous pouvons constater, sur la **Figure B.21**, que les individus déclarés en individus supplémentaires dans l'analyse (voir **Tableau B.5**) sont distinctement éloignés des autres individus. De plus, pour avoir une idée de la valeur des variables correspondant à chaque individu, il suffit de tracer une droite perpendiculaire au vecteur représentant la variable choisie. Par exemple, l'expérience numéro 1 (**Figure B.21**) a été réalisée avec une grande valeur de X1 et de X5 et une petite valeur de X4 et X3.

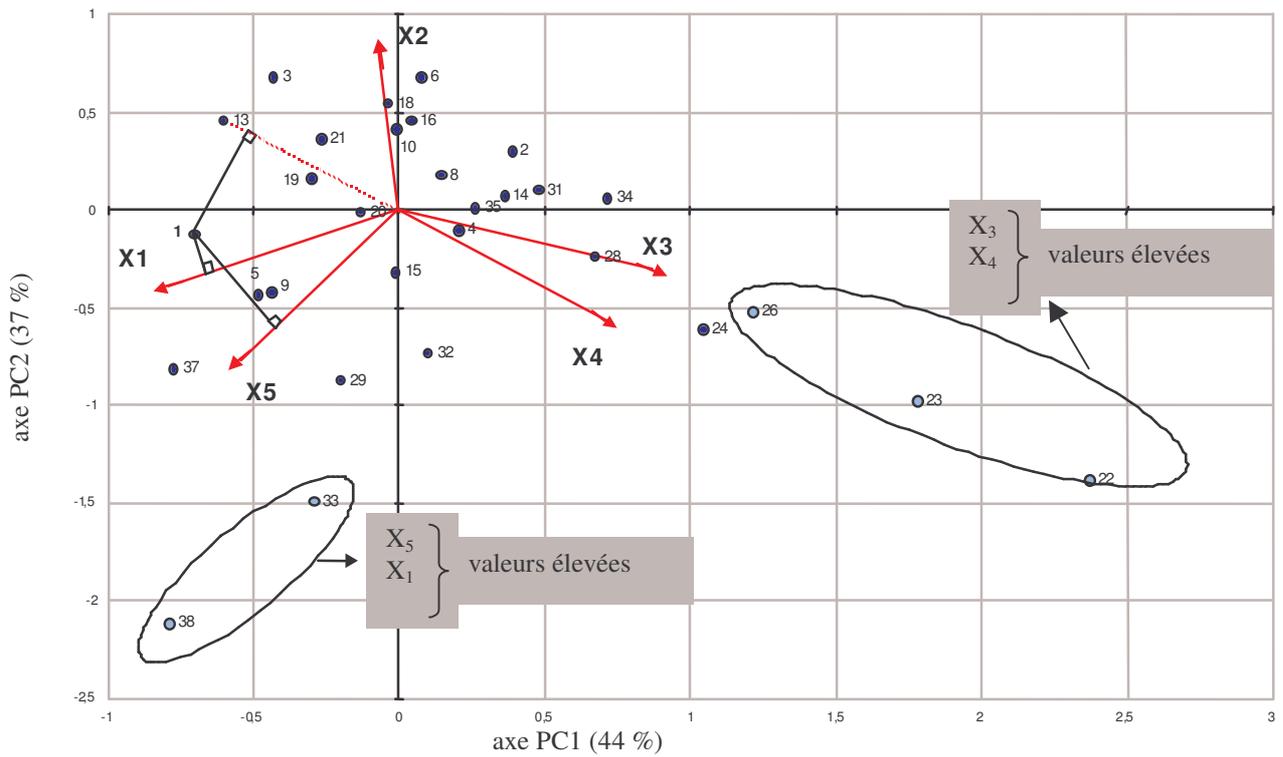


Figure B.21 Extension du Biplot

D'autres analyses complètent celle-ci pour observer les familles d'individus comme, par exemple, la classification ascendante hiérarchique que nous allons décrire dans le chapitre suivant.

Chapitre IV La classification ascendante hiérarchique

(C.A.H)

IV.1. INTRODUCTION

Les méthodes de classification ont pour but de regrouper les individus en un nombre restreint de classes homogènes sous forme d'arbre de classification, appelé couramment **dendrogramme**. Il s'agit ici de décrire les données en procédant à une réduction du nombre d'individus. Il sera question de «classification automatique» et non pas de méthodes subjectives ou visuelles comme l'analyse du Biplot de l'A.C.P. Il existe deux grands types de méthodes de classification [3] :

- les méthodes non hiérarchiques qui produisent directement une partition en un nombre fixé de classes ;
- les méthodes hiérarchiques, auxquelles nous nous intéresserons plus particulièrement dans ce chapitre, qui produisent des suites de partitions en classes de plus en plus vastes à l'image des célèbres classifications des zoologistes en espèces, genres, familles... etc.
- Depuis quelques années, avec le développement des gros calculateurs, d'innombrables algorithmes de classification ont vu le jour. Il n'est pas question de les passer tous en revue dans ce chapitre [15]. Nous allons simplement décrire la méthode qui nous a semblé la plus efficace et la plus utilisée : la classification ascendante hiérarchique (C.A.H.). Puis, nous illustrerons cette méthode à partir d'un exemple simple, la construction d'un dendrogramme et nous terminerons en présentant le dendrogramme obtenu à partir des individus du **Tableau B.2** (chapitre II).

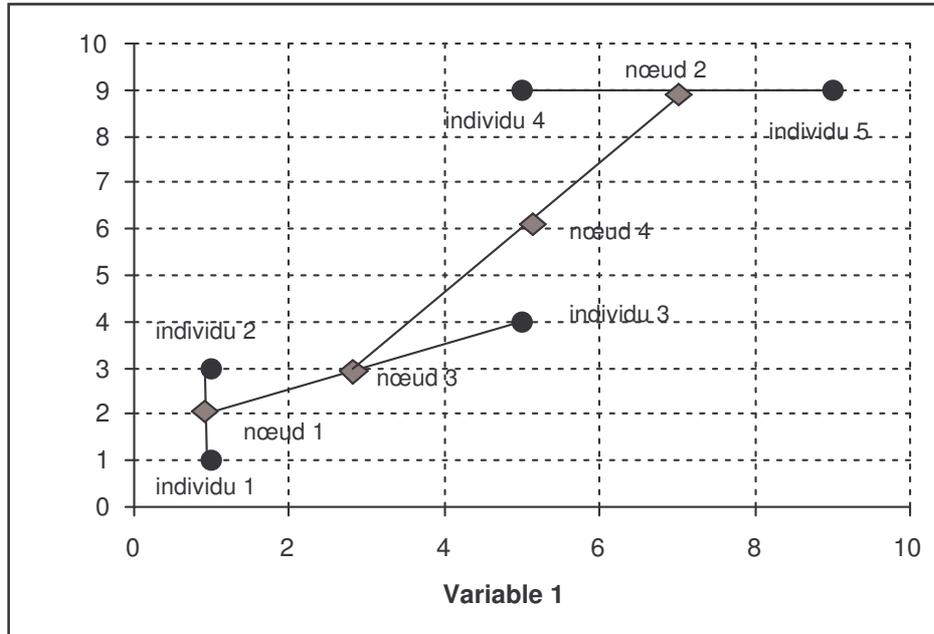
IV.2. PRINCIPE ⁽¹⁾

La classification ascendante hiérarchique vient compléter les analyses précédentes ; elle consiste à obtenir une typologie des individus sous forme de dendrogramme (**Figure B.23**). Pour construire ces arbres de classification, il est fortement conseillé d'utiliser un logiciel d'analyse de données (voir référence des logiciels **Annexe 4**). Pour comprendre le principe de construction d'un dendrogramme, nous allons prendre un exemple simple comportant 5 individus caractérisés par 2 variables. (voir **Tableau B.11**). On peut tracer, à partir du **Tableau B.11**, le nuage de points des individus dans le plan des variables 1 et 2. On obtient ainsi le graphique de la **Figure B.22**.

⁽¹⁾François LOUVET, *Analyse en composantes principales, Notes de cours, ENSCI, 2000*

Tableau B.11 Définition des individus

	Variable 1	Variable 2
Individu 1	1	1
Individu 2	1	3
Individu 3	5	4
Individu 4	5	9
Individu 5	9	9

**Figure B.22** Nuage des individus

La construction d'un arbre de classification (dendrogramme) fait appel à différents principes mathématiques et notamment à la notion de distance. De manière simplifiée, on cherche à construire des nœuds entre les individus. On va détailler la construction des nœuds pour comprendre au mieux le principe de la classification.

Un des objectifs de l'analyse est de décrire les proximités entre les individus c'est-à-dire de rechercher s'il existe des individus semblables et des individus qui se différencient d'autres groupes d'individus semblables. Pour cela, on calcule le carré des distances euclidiennes entre les individus. Les valeurs correspondantes reportées dans le **Tableau B.12**, sont calculées à partir de la formule :

$$d^2(\text{individu } i, \text{individu } j) = (x_j - x_i)^2 + (y_j - y_i)^2$$

avec (x_i, y_i) et (x_j, y_j) coordonnées des individus i et j dans le plan des variables 1 et 2.

Tableau B.12 Carré des distances euclidiennes (construction du noeud 1)

	Individu 1	Individu 2	Individu 3	Individu 4	Individu 5
Individu 1	0				
Individu 2	4	0			
Individu 3	25	17	0		
Individu 4	80	52	25	0	
Individu 5	128	100	41	16	0

Les individus seront d'autant plus proches que leur distance euclidienne sera faible. Ensuite, on remplace les individus les plus proches (dans cet exemple, les individus 1 et 2) par un nouveau point appelé **noeud 1** de coordonnées $\left(\frac{x_1 + x_2}{2}; \frac{y_1 + y_2}{2}\right)$ puis, on calcule à nouveau le carré des distances euclidiennes entre les points restants (**Tableau B.13**).

Tableau B.13 Carré des distances euclidiennes (construction du noeud 2)

	Individu 3	Individu 4	Individu 5	Nœud 1
Individu 3	0			
Individu 4	25	0		
Individu 5	41	16	0	
Nœud 1	20	65	113	0

Les individus 4 et 5 sont à présent les plus proches, on les remplace par le noeud 2 (voir **Figure B.22**) et on calcule à nouveau la distance entre les points restants (**Tableau B.14**).

Tableau B.14 Carré des distances euclidiennes (construction du noeud 3)

	Individu 3	Nœud 1	Nœud 2
Individu 3	0		
Nœud 1	20	0	
Nœud 2	29	85	0

L'individu 3 et le noeud 1 sont les plus proches. On obtient donc le noeud 3.

Tableau B.15 Carré des distances euclidiennes (construction du noeud 4)

	Nœud 2	Nœud 3
Nœud 2	0	
Nœud 3	52	0

Ainsi le dernier noeud (noeud 4) sera défini grâce aux noeuds 2 et 3.

Il ne reste plus alors qu'à représenter le résultat des étapes précédentes sous forme d'un arbre de classification (**Figure B.23**) également appelé dendrogramme.

Un vocabulaire assez précis est utilisé dans le cadre de la Classification Ascendante Hiérarchique :

- les individus forment des classes à un seul élément, appelées *classes terminales* (c désigne le nombre de classes terminales),
- le regroupement de 2 classes forme un *nœud*. Le premier de ces nœuds a pour *numéro* le nombre total de classes terminales +1 ($c+1$). Chaque regroupement donne lieu à un nouveau nœud dont le numéro suit exactement celui du regroupement précédent. Le dernier nœud aura donc pour numéro: $2.c - 1$. Par extension, on désignera également les individus par un numéro de nœud compris entre 1 et c . Chaque nœud est ainsi formé de 2 classes que l'on appelle ses *successeurs*, l'un étant appelé *l'aîné* et l'autre le *benjamin* (de façon arbitraire). L'*indice de niveau* d'un nœud est défini par la distance entre les 2 classes formant ce nœud. Il permet de mesurer et de représenter l'importance de la différence entre ces 2 classes.

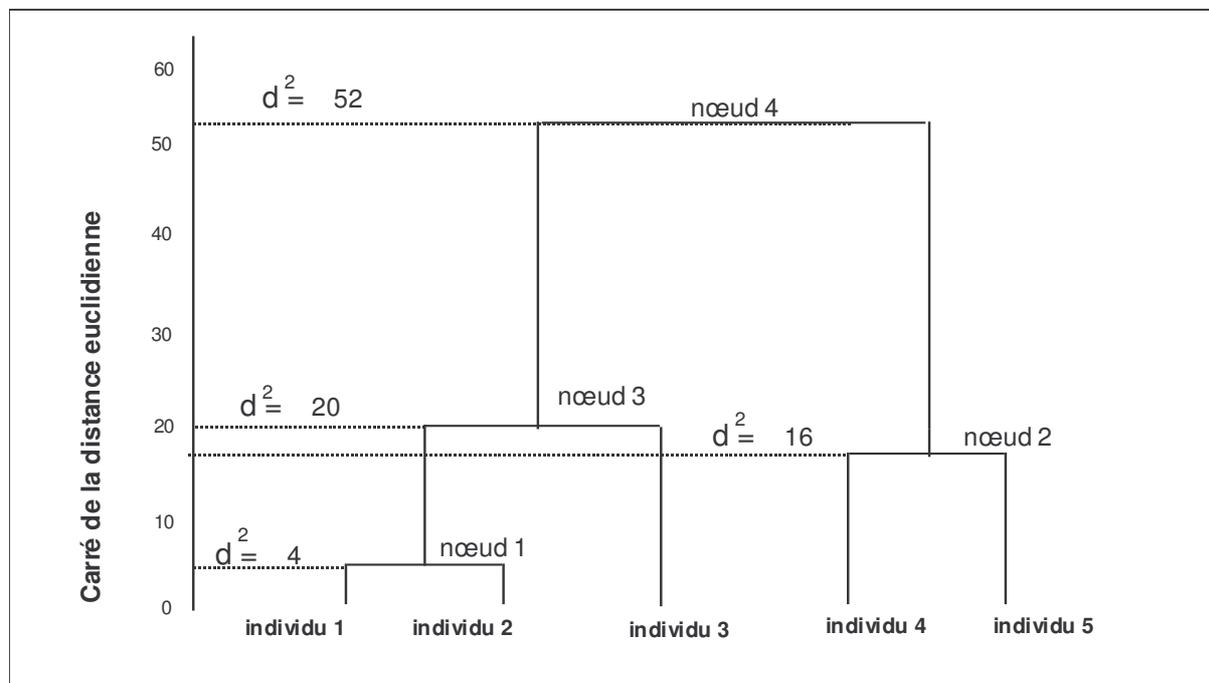


Figure B.23 Arbre de classification/Dendrogramme

Cette procédure est basée sur 2 choix : la détermination d'un **critère de ressemblance** entre les individus et la détermination d'une distance entre classes, procédé appelé **critère d'agrégation**.

□ LA DETERMINATION D'UN CRITERE DE RESSEMBLANCE ENTRE LES INDIVIDUS :

Il existe plusieurs critères mathématiques et statistiques de similarités/dissimilarités. Ces critères sont répertoriés dans le **Tableau B.16** et sont expliqués dans la bibliographie [16].

Le critère de ressemblance entre les individus le plus utilisé est la **distance euclidienne**.

□ **LA DETERMINATION D'UNE DISTANCE ENTRE CLASSES : PROCEDE APPELE CRITERE D'AGREGATION :**

Plusieurs méthodes d'agrégation sont disponibles : **Méthode de Ward** (inertie), **Méthode de Ward** (variance), Lien complet, Lien simple, Lien fort, Lien flexible, Lien moyen, Lien proportionnel ;

La difficulté du choix du critère d'agrégation réside dans le fait que ces critères peuvent déboucher sur des résultats différents. Le critère le plus couramment utilisé est celui du moment centré d'ordre 2, appliqué à une matrice centrée - réduite (Méthode de Ward). Des exemples de calcul selon les différents critères ont été expliqués par **Lena Sanders [16]**.

Tableau B.16 Critère de ressemblance entre les individus

Similarité	Dissimilarité
Corrélation de Pearson	Distance euclidienne
Corrélation de Spearman	Distance du khi ²
Corrélation de Kendall	Distance de Manhattan
Inertie	Dissimilarité de Pearson
Covariance (n)	Dissimilarité de Spearman
Covariance (n-1)	Dissimilarité de Kendall
Similarité générale	Dissimilarité générale

IV.3. ARBRE DE CLASSIFICATION

Cette représentation de la typologie des individus complète généralement les résultats d'une analyse en composantes principales, elle est plus précise et plus fiable pour la construction des familles d'individus.

A titre d'exemple, la **Figure B.24** représente l'arbre de classification obtenu à partir des individus contenus dans le **Tableau B.2**.

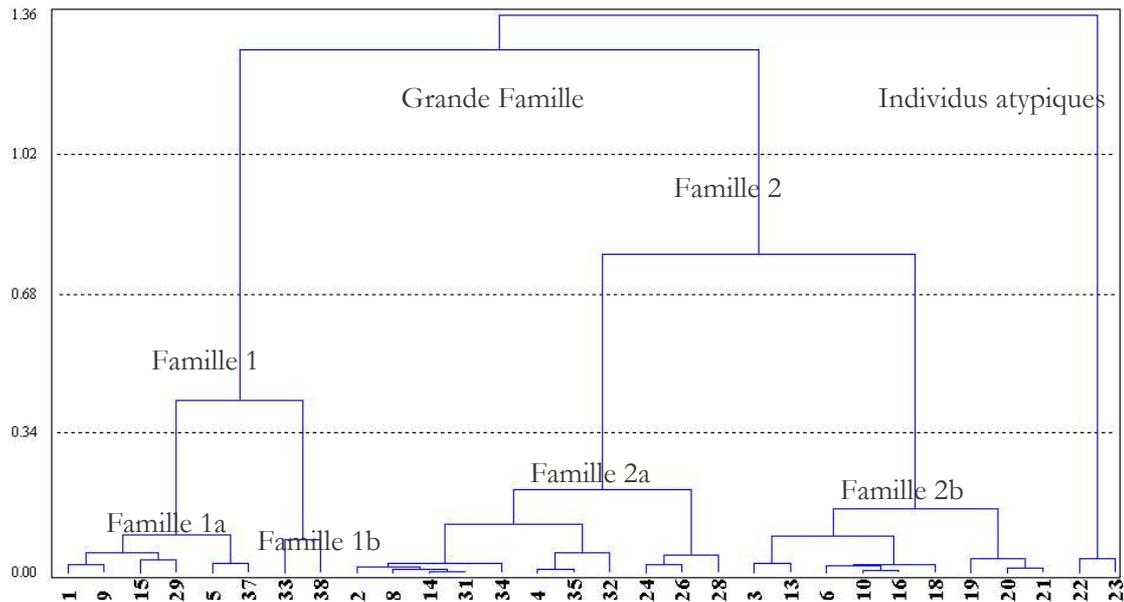


Figure B.24 Dendrogramme des individus du **Tableau B.2**

On peut interpréter ce dendrogramme de la façon suivante :

- on obtient un grand groupe (une grande famille et un groupe de deux individus atypiques (individus 22 et 23). On peut en conclure que ces deux expériences sont très différentes des autres. Notons que nous avons déjà cette information en regardant le biplot de l'A.C.P. (**Figure B.21**) mais ici nous voyons de manière plus claire que ces deux individus sont vraiment en marge de tous les autres,
- dans la grande famille d'individus, on observe deux sous familles : la famille 1 et la famille 2. Dans ces sous-familles, ils existent encore des petites familles (familles 1a et 1b, famille 2a et 2b).
- A partir de ce dendrogramme, un certain nombre de conclusions pourront être déduites par l'expérimentateur :
- quelles sont les expériences les plus semblables c'est à dire celles qui ont à peu près les mêmes valeurs des variables ?
- quels sont les groupes d'expériences ?
- quelles sont les différences entre ces différents groupes ?
- quelles sont les variabilités des réponses obtenues ?

Dans la partie suivante, nous allons aborder une autre méthode d'analyse de données : l'Analyse Factorielle Discriminante (A.F.D.).

Chapitre V L'Analyse Factorielle Discriminante (A.F.D.)

V.1. INTRODUCTION

L'A.F.D. est une analyse factorielle qui s'intéresse aux relations entre **une variable qualitative** et un **ensemble de variables quantitatives** [1]. Comme de nombreuses analyses de données, l'analyse factorielle discriminante peut être, simplement descriptive mais également prédictive.

Considérons un ensemble d'individus sur lequel on observe un caractère qualitatif prenant **q** modalités. Chaque individu étant repéré par une seule modalité de ce caractère, on définit une partition de l'ensemble des individus en **q** classes disjointes. Par ailleurs, on mesure sur les mêmes individus **p** caractères quantitatifs.

En fait, l'A.F.D. recouvre deux aspects, elle permet :

- de séparer au mieux l'ensemble des **q** classes à l'aide de **p** variables (on dit que l'analyse discriminante permet la discrimination de populations),
- de classer les observations. Pour chaque observation (ou expériences), on pourra définir une probabilité d'appartenir à chacune des populations de la variable qualitative,
- d'estimer un taux de mauvais classement.

Cette opération de classement peut aussi être faite pour des observations nouvelles, qui n'ont pas participé à la discrimination. L'A.F.D. est un outil important dans ce que l'on appelle l'aide à la décision. Cette analyse se base sur de nombreux principes statistiques que nous n'avons pas détaillés dans ce chapitre (analyse de la variance, distance de Mahalanobis...) [17][11].

V.2. PRINCIPE

On considère ici 10 individus ($n = 10$) représentés par 2 variables quantitatives (X_1 et X_2). La variable X_1 prend des valeurs comprises entre 0 et 13 et X_2 de 0 à 9. Chaque individu appartient soit à la classe A (ou groupe A) soit à la classe B (groupe B) de la variable qualitative Y.

Ces données sont répertoriées dans le **Tableau B.17** ci-dessous.

Tableau B.17 Tableau de données brutes

	X1	X2	Y
1	0	3	A
2	2	1	A
3	4	5	A
4	6	9	A
5	8	7	A
6	5	2	B
7	7	0	B
8	9	4	B
9	11	8	B
10	13	6	B

Supposons que l'on ait 3 expériences supplémentaires à réaliser (voir **Tableau B.18**) et que l'on souhaite connaître, avant de réaliser ces expériences, quelle serait la valeur qualitative de Y (A ou B) prise par ces expériences supplémentaires SUP1, SUP2 et SUP3.

Tableau B.18 Expériences supplémentaires

Expérience	X1	X2
SUP1	5	3
SUP2	2	14
SUP3	6	8

Dans le **Tableau B.19**, sont reportés les résultats de l'Analyse Factorielle Discriminante. On peut voir, que l'A.F.D. a classé les individus supplémentaires dans le groupe A avec une probabilité égale à 0,58 pour l'expérience SUP1, 1 pour SUP2 et 0,99 pour SUP3.

Tableau B.19 Classement des individus supplémentaires

Individu	Groupe	Prob. A	Prob. B
SUP1	A	0,583	0,417
SUP2	A	1,000	0,000
SUP3	A	0,997	0,003

Dans la plupart des cas, le tableau de données à analyser est beaucoup plus important que celui ci et il est alors intéressant de visualiser les résultats sur un graphique au lieu d'un autre tableau.

Nous allons voir dans le paragraphe V.2.1 de ce chapitre, un autre exemple d'application de l'A.F.D. plus complexe.

V.2.1. APPLICATION DE L'A.F.D

Dans ce paragraphe, nous allons montrer en quoi cette méthode nous a été utile pour projeter des expériences à venir compte tenu de celles déjà réalisées.

Les expériences déjà réalisées, sont reportées dans le **Tableau B.20**. Elles regroupent 55 individus caractérisés par 5 variables quantitatives (X1, X2, X3, X4 et X5) et une variable qualitative Y. La variable Y peut prendre 3 valeurs qualitatives : groupes A, B ou C.

Pour les treize expériences à venir (individus Sup1 à Sup13) reportées dans le **Tableau B.21**, on veut connaître la valeur prise par la variable Y (A, B ou C). Pour cela, on réalise une analyse factorielle discriminante à partir des données du **Tableau B.20** auquel on ajoute les treize expériences du **Tableau B.21** déclarées en tant qu'individus supplémentaires.

Tableau B.20 Individus de l'A.F.D.

	X1	X2	X3	X4	X5	Y
78	0,5	34	850	2,2	660	A
57	0,3	60	810	1,92	1190	A
72	0,3	30	850	1,9	960	A
54	0,3	60	795	1,76	1170	A
55	0,3	65	790	1,73	1300	A
58	0,3	60	805	1,7	770	A
86	0,8	35	850	1,654	790	A
59	0,3	60	820	1,53	870	A
87	0,5	30	850	1,511	860	A
56	0,3	60	800	1,51	1340	A
38	0,8	45	800	1,5	2400	A
85	0,8	30	850	1,36	680	A
89	0,8	26	850	0,909	550	A
75	0,3	6	850	0,869	660	A
88	0,8	23	850	0,8	420	A
51	0,3	30	790	0,7	640	A
69	0,3	6	850	0,58	740	A
76	0,3	8	850	0,37	610	A
73	0,3	20	850	1,78	980	B
45	0,5	26	800	1,157	900	B
47	0,5	29	800	0,97	1200	B
37	0,8	30	800	0,87	290	B
48	0,5	27	800	0,8	870	B
46	0,5	23	800	0,7	950	B
43	0,5	21	800	0,63	950	B
83	0,8	16	850	0,61	690	B
52	0,3	20	800	0,56	1730	B
71	0,3	8	850	0,48	780	B
53	0,3	18	820	0,48	690	B
42	0,5	12	800	0,37	1300	B
68	0,3	4	850	0,34	580	B
44	0,5	6	800	0,28	860	B
70	0,3	1	850	0,2	440	B
84	0,8	20	850	0,46	360	B
67	0,3	2	850	0,39	470	B
11	0,8	15	500	0,42	472	C
12	0,8	15	600	0,44	264	C
13	0,8	15	700	0,33	133	C
14	0,8	15	800	0,81	370	C
10	0,8	15	400	1,1	400	C
32	0,8	30	850	0,8	300	C
33	0,8	30	850	1,4	640	C
34	0,8	30	825	0,8	580	C
35	0,8	30	825	0,7	490	C
36	0,8	30	800	0,7	870	C
19	0,8	30	850	0,24	2500	C
50	0,3	15	790	0,56	1030	C
62	0,3	15	850	0,59	770	C
63	0,3	17	850	0,79	420	C
64	0,3	15	820	0,65	830	C
31	0,8	30	850	0,5	520	C
28	0,8	30	850	0,48	820	C
29	0,8	30	850	0,4	330	C
90	0,5	2	850	0,12	440	C
30	0,8	30	850	0,35	230	C

Tableau B.21 Individus supplémentaires dans l'A.F.D.

	X1	X2	X3	X4	X5
Sup1	0,3	12	795	0,37	770
Sup2	0,5	4	790	0,34	790
Sup3	0,3	6	805	0,28	870
Sup4	0,5	1	850	0,2	860
Sup5	0,3	20	820	0,46	1340
Sup6	0,8	2	850	0,39	2400
Sup7	0,3	15	800	0,42	680
Sup8	0,3	15	800	0,44	550
Sup9	0,3	15	850	0,33	660
Sup10	0,8	15	850	0,81	420
Sup11	0,8	15	850	0,79	640
Sup12	0,8	15	850	0,65	330
Sup13	0,5	17	790	0,5	440

Les résultats obtenus, reportés dans le **Tableau B.22**, montrent que les individus Sup1 à Sup9 ont été classés dans le groupe B et les individus Sup10 à Sup13 dans le groupe C. Ce tableau regroupe également les coordonnées de ces individus sur les nouveaux axes de projection F1 et F2 au même titre que dans une Analyse en Composante Principale. Grâce à ces coordonnées, il est possible de faire une représentation graphique (**Figure B.25**).

Tableau B.22 Classement, probabilités d'appartenance et coordonnées des individus supplémentaires

Individu	Groupe	Prob. A	Prob. B	Prob. C	F1	F2
Sup1	B	0,045	0,740	0,215	3,453	-1,804
Sup2	B	0,019	0,687	0,294	2,872	-1,751
Sup3	B	0,025	0,818	0,158	3,248	-2,224
Sup4	B	0,013	0,771	0,216	2,772	-2,134
Sup5	B	0,074	0,784	0,141	3,892	-1,996
Sup6	B	0,007	0,885	0,108	2,758	-2,806
Sup7	B	0,064	0,693	0,243	3,600	-1,589
Sup8	B	0,072	0,663	0,265	3,628	-1,472
Sup9	B	0,058	0,723	0,219	3,588	-1,711
Sup10	C	0,099	0,345	0,556	3,571	-0,380
Sup11	C	0,093	0,400	0,508	3,554	-0,578
Sup12	C	0,055	0,293	0,653	3,202	-0,283
Sup13	C	0,059	0,442	0,499	3,304	-0,782

Sur cette figure, les triangles représentent les individus supplémentaires dont on recherche l'appartenance au groupe A, B ou C. Ces groupes sont représentés par différentes couleurs : A (bleue),

B (rose) et C (vert). Les triangles roses représentent donc les individus supplémentaires classés dans le groupe B et les triangles verts, ceux classés dans le groupe C.

On constate également, conformément au **Tableau B.22**, qu'aucune des 13 expériences supplémentaires n'a été classée dans le groupe A (triangle bleu).

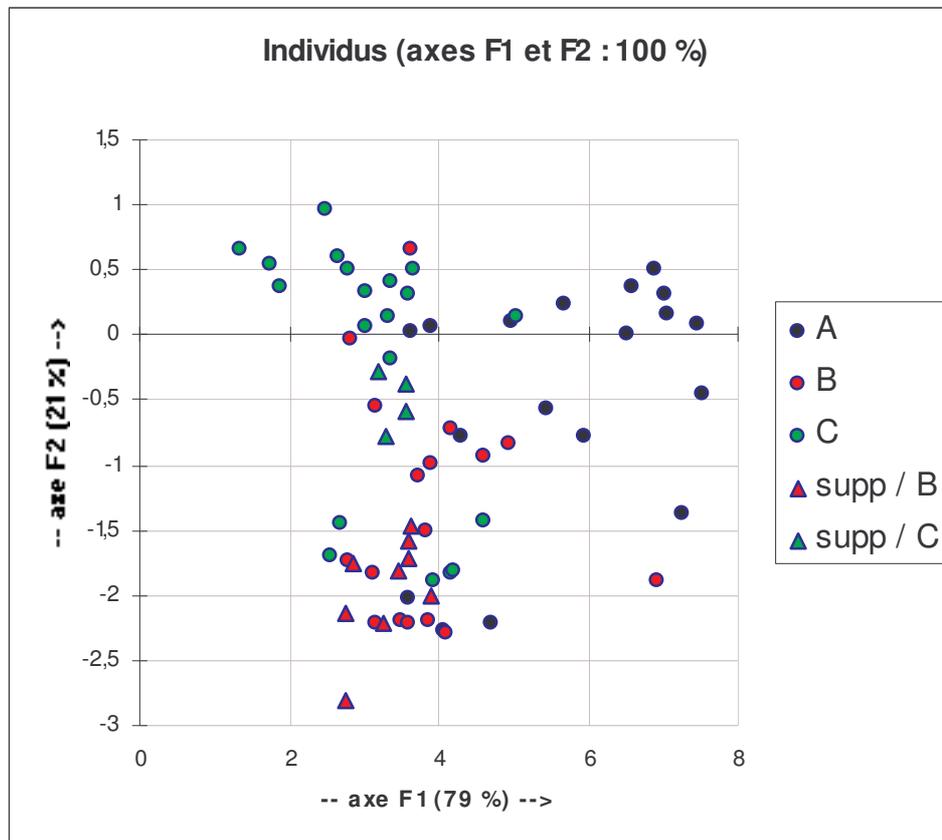


Figure B.25 Classement des individus dans les groupes de Y

L'A.F.D. est une méthode très pertinente qui permet de prédire la suite des expériences. Nous verrons dans la partie C de ce manuscrit son application à l'optimisation des couches minces de titanate de baryum et de strontium obtenues par ablation laser.

Conclusion sur l'analyse de données

Statistique descriptive perfectionnée, l'analyse de données est la description d'ensembles de données de taille importante et la description des relations entre plusieurs ensembles de données. Le but est généralement de fournir, à partir d'un grand tableau de données, un résumé en fonction d'un critère précis, en vue d'une action. C'est un outil précieux pour le chercheur qui veut extraire le maximum d'informations des données collectées.

Dans cette partie, nous avons tout d'abord montré l'intérêt des outils classiques de la statistique descriptive (paramètres de dispersion, variances, corrélation, ...). L'utilisation de diagrammes et de graphiques constitue un moyen commode pour organiser et présenter des données et saisir beaucoup d'informations d'un seul coup d'œil. Par exemple, on peut se servir de diagrammes à boîtes et moustaches pour récapituler des données présentées dans un tableau mais aussi pour déceler des erreurs de saisies ou des individus atypiques.

La première méthode d'analyse multidimensionnelle de données exposée dans cette partie, est l'analyse en composantes principales qui s'applique à un tableau de données quantitatives croisant individus et variables. Nous avons mis en évidence des composantes orthogonales (c'est-à-dire indépendantes) tenant compte de la variance totale de toutes les variables étudiées. Les représentations graphiques de l'A.C.P. (cercle des corrélations, biplot) permettent par une simple lecture graphique de saisir les relations linéaires entre individus et variables.

Nous avons montré que la classification ascendante hiérarchique consiste à découper une population d'individus en différentes classes tenant compte des variables qui les caractérisent et à étudier les ressemblances entre les variables à travers les individus.

L'analyse factorielle discriminante est la dernière méthode présentée. Contrairement à l'A.C.P., cette méthode permet d'établir des relations entre une variable qualitative et un groupe de variables quantitatives. En particulier, nous avons vu qu'à partir d'individus caractérisés par un ensemble de variables quantitatives partitionnées en groupes, il est possible de classer un nouvel individu dans un groupe donné lorsque l'on connaît les valeurs de l'ensemble des variables prises par cet individu.

Cependant, toutes ces méthodes menées à leur terme appellent à l'explication et nécessitent la réflexion de l'expérimentateur. Nous verrons lors de l'étude d'exemples (partie C) que l'analyse de données conduit l'expert à se poser des questions car elle ne fournit pas, à elle seule, les réponses adéquates.

RAPPEL : NOTATIONS ANALYSES DE DONNEES

p : nombre de variables

n : nombre d'individus

x_{ij} : élément courant du tableau brut, $j^{\text{ème}}$ variable pour le $i^{\text{ème}}$ individu

\bar{X}_j : moyenne arithmétique des valeurs prises par la variable x_j

X_{ij} : données centrées réduites

$\sigma(x_j)$: écart type de la variable x_j

Var (X_j) : variance de la variable X_j

Q : fonction quantile

F(x) : fonction de répartition

I_a : intervalle interquantile d'ordre a

$Q_{75}-Q_{25}$: écart interquartile

$Q_{90}-Q_{10}$: écart interdécile

$Q_{99}-Q_1$: écart intercentile

$y_{ij} = ax_{ij} + b + e_{ij}$ avec $i = 1, \dots, n$

e_{ij} : un terme résiduel.

x_j et y_j , deux variables

Cov (x,y) : covariance entre 2 variables x et y

R(X,Y) : coefficient de corrélation entre X et Y

R²(x,y) : coefficient de détermination entre x et y

PC1 (Principal Component 1) \cong Axe 1 \cong Z1 : 1^{er} axe principal

Z_{ij} : coordonnées des individus dans le repère de l'A.C.P.

λ_i : valeur propre de la $i^{\text{ème}}$ variable

[Z] : matrice des nouvelles coordonnées de l'A.C.P.

[V] : matrice des vecteurs propres

[X] : matrice des coordonnées centrées réduites

CTR (i, axe j) : contribution des individus i à la construction d'un axe j

cos² (α_i) : qualité de représentation d'un individu i

α_i : angle entre l'axe principale et l'individu i

c : nombre total de classes terminales

Bibliographie Partie B

- [1] H.HOTTELING, *Analysis of complex of statistical variables into principal components*, Journal of Educational Psychology, 1933, vol. 24, 417 – 441, 408 –520.
- [2] C. SPEARMAN, *General intelligence objectively determined and measured*, American Journal of Psychology, 1904, vol. 28, 129-149.
- [3] J-M BOUROCHE, G. SAPORTA, *L'analyse des données*, Ed. Que sais-je ?, 2002, 3-4.
- [4] M. JAMBU, *Méthodes de base de l'analyse des données*, Collection technique et scientifique des télécommunications, Ed. Eyrolles, 1999, XI-5.
- [5] H. HOTELLING, *Relations between two sets of variables*, Biometrika, 1936, vol.28, 129-149.
- [6] K.PEARSON, *On lines and planes of closest fit to system of points in space*, Phil.Mag., 1901,n° 11, 559-572.
- [7] M. VOLLE, *Analyse des données*, Collection « économie et statistiques avancées », Ed. Economica, 1993.
- [8] P. CASIN, *Analyse des données et des panels de données*, Ed. DeBoeck, Université perspectives marketing, 1999.
- [9] G. CHAUVAT, J-PREAU, *Statistiques descriptives*, Ed Armand colin Editeur, 2002.
- [10] J. DE LAGARDE, *Initiation à l'analyse des données*, Ed. Dunod.
- [11] G.SAPORTA, *Probabilités analyse des données et statistique*, Ed. Technip, 1990.
- [12] G.PHILLIPPEAU, *Comment interpréter les résultats d'une analyse en composantes principales ?* Collection STAT-ITCF-Ed. 1992.
- [13] .P FENELON., *Qu'est ce que l'analyse des données ?* , Ed. Lefonen , PARIS, 1981.
- [14] R. RODRIGUEZ HERRERRA, D. SALLES-LE GAC, *Initiation à l'analyse factorielle des données*, Ellipse, 2002.
- [15] F.CAILLIEZ, J.P. PAGES, *Introduction à l'analyse des données*, Société de Mathématique Appliquée et de Sciences Humaine, Ed. Scientifique, 1976.
- [16] L. SANDERS, *L'analyse statistique des données en géographie*, Collection Alidade. Groupement d'Intérêt Public Reclus. Montpellier. 1989.
- [17] R. TOMASSONE, *Comment interpréter les résultats d'une analyse factorielle discriminante ?*, Collection STAT-ITCF-Ed. 1992.

Partie C :

**Exemples d'applications de l'analyse de
données et des plans d'expériences**

Chapitre I :

Optimisation des couches minces d'alumine

réalisées par PECVD

Dans le cadre d'une collaboration avec le Laboratoire des Sciences des Procédés Céramiques et Traitements de Surface (**S.P.C.T.S.**), nous avons été amenés à optimiser un bâti de dépôt par P.E.C.V.D. (technique de Dépôt Chimique en Phase Vapeur assistée par Plasma). Toutes les expériences mentionnées dans ce chapitre, ont été réalisées par **H. Hidalgo [1]** et **T. Haure [2]** doctorants au S.P.C.T.S.

I.1. FORMALISATION DU PROBLEME

I.1.1. OBJECTIF

Avant de résumer les résultats obtenus à partir de l'ensemble des méthodes et outils utilisés, une brève description du dispositif expérimental et des principaux paramètres de dépôt intervenant dans cette étude, va être présentée.

I.1.1.a. Présentation de l'étude

La méthode CVD (dépôt chimique en phase vapeur) permet de réaliser des couches sur un substrat chauffé à partir d'un précurseur gazeux et d'une réaction chimique. Pour que les réactions chimiques aient lieu, le substrat doit être porté à une température qui varie, selon les matériaux à déposer, entre 500 °C et 2000 °C. A ces températures, de nombreux substrats peuvent se détériorer ou subir d'importantes modifications de caractéristiques. Afin de diminuer considérablement la température des substrats (en général inférieure à 300 °C), on utilise la méthode PECVD dans laquelle l'énergie thermique nécessaire aux réactions est remplacée par l'action des électrons énergétiques créés dans une décharge électrique. Il existe deux types de décharges électriques : les plasmas thermiques tels que des arcs électriques à la pression atmosphérique et les plasmas froids tels que des décharges entretenues à basse pression. La technique de dépôt PECVD développée au SPCTS est basée sur ce deuxième type de décharge.

I.1.1.b. Dispositif expérimental

Le dispositif expérimental de PECVD est présenté sur la **Figure C.1**. L'objectif ici n'est pas de décrire les différents éléments qui le constituent mais de faire ressortir les principaux paramètres qui vont intervenir au cours des différentes étapes d'optimisation.

Dans cette étude, le réacteur de dépôt est utilisé pour réaliser des couches d'alumine sur substrat silicium. Le protocole de dépôt et les outils de caractérisation des films font intervenir de nombreux paramètres. On rappelle que les variables d'entrée sont les paramètres de dépôt et les variables de sortie, les mesures réalisées sur les films d'alumine après dépôt.

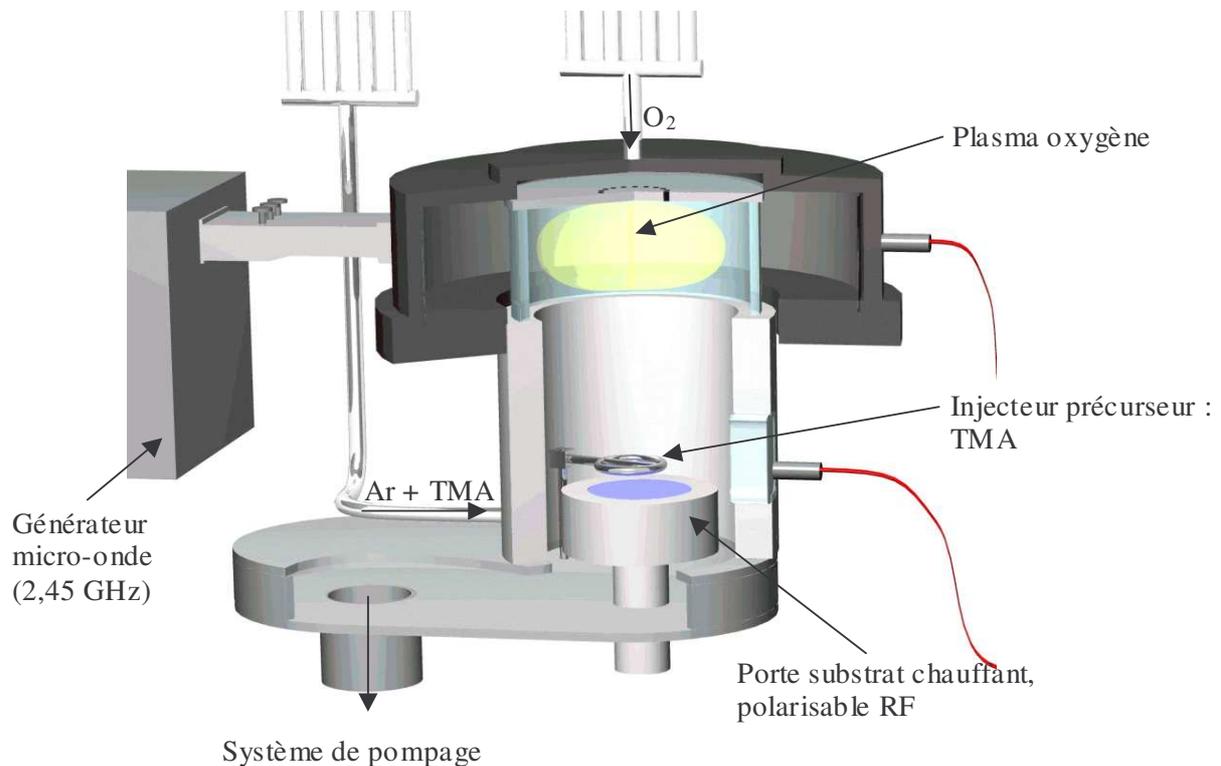


Figure C.1 Réacteur de dépôt PECVD

Le temps de dépôt est fixé à 10 minutes.

Les objectifs des différentes analyses que nous allons développer dans la suite de ce chapitre sont :

- de calculer les effets des variables d'entrées sur la fluctuation des variables de sortie, de détecter les facteurs les plus influents. Le travail réalisé par **H. Hidalgo** en collaboration avec **F. Louvet**, dans le cadre de son travail de thèse a permis d'identifier les paramètres prépondérants. Nous avons utilisé ces résultats pour expliquer la construction d'un **plan de criblage** ainsi que les différentes analyses mathématiques et statistiques nécessaires à ce type de plan,

- de corrélérer les facteurs de sorties entre eux (prise de masse, vitesse d'attaque, densité, adhérence, OH/e) afin d'aider l'expérimentateur à comprendre la structure des films réalisés (**A**nalyse en **C**omposantes **P**incipales),
- de corrélérer les propriétés des films aux données du plasma (spectrométrie d'émission optique) afin d'identifier le chemin réactionnel menant à la construction des dépôts (**A.C.P**),
- de corrélérer les propriétés des films à la morphologie des splats obtenus par projection plasma (**A.C.P**),
- d'optimiser l'adhérence des films (validation du bâti de multifissuration et comparaison aux propriétés des couches obtenues par les autres techniques) grâce à un **plan à surface de réponse**.

Objectif : « Dégager les paramètres ayant une forte influence sur les propriétés des dépôts pour ensuite mener une étude plus complète de l'influence de ces paramètres sur les propriétés des films et des espèces dans le plasma »

I.1.2. VARIABLES DE SORTIE

Les sorties peuvent être classées en **cinq grands types** : les vitesses de dépôt, les marqueurs de qualité du film, l'homogénéité en épaisseur du film sur un substrat de grand diamètre, les propriétés mécaniques des couches et la composition du plasma.

I.1.2.a. Vitesses de dépôt

- **dm/St** ($\mu\text{g}\cdot\text{cm}^{-2}\cdot\text{min}^{-1}$) : la **prise de masse** qui caractérise la quantité de matière déposée par unité de surface et de temps sur le substrat. Elle représente la différence de masse entre l'échantillon pesé avant et après dépôt.
- **V_{dépôt}** ($\mu\text{m}\cdot\text{min}^{-1}$) : la **vitesse de dépôt** du film c'est-à-dire l'épaisseur déposée par unité de temps.

I.1.2.b. Marqueurs de qualité du film

- **Densité** : la densité est obtenue à partir de la prise de masse et du volume du dépôt. Le volume est déterminé à partir de l'épaisseur mesurée après attaque à l'acide fluorhydrique mais, en 1 seul point, d'où une incertitude importante car les dépôts ne sont pas forcément homogènes en épaisseur.

- V_{dissol} ($\text{nm}\cdot\text{s}^{-1}$) : la **vitesse de dissolution** appelée aussi **vitesse d'attaque**. Les dépôts d'alumine peuvent être dissous par immersion pendant un temps donné (1 min) dans une solution aqueuse d'acide fluorhydrique. La mesure de la hauteur de la marche (épaisseur dissoute) peut fournir des informations sur la qualité des couches. La vitesse de dissolution doit être la **plus faible** possible.
- **OH/e** (u.a) : représente le **taux d'impuretés**. Les impuretés sous forme OH présentes dans la couche sont déterminées par spectrométrie infrarouge **OH/e** doit être le **plus faible** possible.

I.1.2.c. Homogénéité en épaisseur du film

La mesure de l'épaisseur e (en μm) du film en plusieurs points, est réalisée par mesure de marches issues du quadrillage du wafer. Les coefficients **H%** et **C%** peuvent être utilisés pour estimer l'homogénéité en épaisseur du film :

$$H\% = \frac{e_{\text{max}} - e_{\text{min}}}{e_{\text{max}} + e_{\text{min}}} = \frac{e_{\text{max}} - e_{\text{min}}}{2 \cdot e_{\text{moy}}} \quad C\% = \frac{\sigma(e)}{e_{\text{moy}}} = \frac{\sqrt{V(e)}}{e_{\text{moy}}}$$

avec e_{max} , e_{min} , e_{moy} : épaisseur maximale, minimale et moyenne

$\sigma(e)$ et $V(e)$: écart type et variance de l'épaisseur.

Plus les coefficients **H%** et **C%** sont faibles, plus la répartition en épaisseur du dépôt est homogène sur le wafer. Pour plus de précision à ce sujet, il faut se reporter à la thèse d'**H. Hidalgo** [1].

I.1.2.d. Propriétés mécaniques

Adhérence : la détermination de l'adhérence entre le film et le substrat, consiste à mesurer la force minimale qui doit lui être appliquée pour permettre la rupture des liaisons présentes entre le dépôt et le substrat ou entre les deux matériaux déposés. Pour cette mesure, la technique utilisée est le test de rayure [2].

I.1.2.e. Composition du plasma

La composition de la phase réactive peut être déterminée par spectrométrie d'émission optique à l'aide d'une fibre optique positionnée entre l'injecteur et le substrat. Les photons émis lors de la désexcitation des espèces dans le plasma sont caractéristiques d'un atome ou d'une molécule simple. Cependant, l'intensité de la raie n'est pas directement proportionnelle à la concentration en espèces dans le plasma.

A l'aide d'un gaz rare, la technique de l'actinométrie permet, à partir du rapport d'intensité entre l'espèce considérée et le gaz rare, de calculer l'évolution du rapport de concentration de l'espèce considérée par rapport au gaz rare. Dans notre cas, six espèces ont été étudiées : **OH**, **AlO**, **O₂⁺**, **Al**, **O** et **H**.

I.1.3. VARIABLES D'ENTREE

Dans cette procédure, les paramètres que l'expérimentateur a choisi de faire varier sont :

- **V_{pompe}** (krpm ou 1000 tr.min⁻¹) : vitesse de rotation de la pompe,
- **P** (Pa) : pression dans l'enceinte, contrôlée par la vitesse de rotation de la pompe turbomoléculaire et le débit de gaz,
- **d(O₂)** (sccm) : débit d'oxygène,
- **d(TMA)** (sccm) : débit de TMA (TriMéthylAluminium),
- **Puiss. MW**(W) : puissance micro-onde,
- **Prép** : mode de préparation de la surface du substrat avant dépôt (présence ou non d'un plasma d'argon (50 sccm, 1 Pa, 1000 W, -500 V, 10 min)),
- **Dist** (cm) : distance entre l'injecteur de TMA et le porte-substrat,
- **Pola** (V) : valeur absolue de la polarisation radiofréquence du porte-substrat (dans la réalité, les valeurs sont négatives).
- **Temp** (°C) : température de consigne du porte-substrat.

I.2. CRIBLAGE DES FACTEURS

Des études préalables réalisées par **H. Hidalgo [1]** en collaboration avec **F. Louvet**, ont permis d'étudier le poids des facteurs grâce à un plan d'expériences appelé **plan de criblage**.

I.2.1. FACTEURS - MODALITES - DOMAINE EXPERIMENTAL

L'influence de la vitesse de pompe, du débit d'oxygène, du débit de TMA, de la température, de la puissance micro-onde, de la polarisation, de la distance injecteur/substrat ainsi que du mode de préparation de la surface du substrat avant dépôt, a été menée via un plan de criblage qui nécessitait 2 valeurs par facteur. Le temps de dépôt a été fixé à 10 minutes pour toutes les expériences. Donc :

Nombre de facteurs : $k = 8$

Nombre de modalités : $m_i = 2$

$N_{\text{total}} = 2^8 = 256$ expériences possibles

Les plans d'expériences que nous allons décrire dans cette partie, vont évidemment montrer qu'il n'est pas nécessaire de réaliser les 256 expériences.

Dans le cas présent, tous les facteurs présentent deux modalités, la première modalité indique conventionnellement l'état de référence. Ces modalités sont reportées dans le **Tableau C.1**.

Tableau C.1 Modalités des facteurs

Facteur	Modalité 1 (-)	Modalité 2 (+)
Prép	plasma	sans plasma
Temp	100°C	800°C
Dist	4 cm	8 cm
Puiss. MW	1000 W	1600 W
V _{pompe}	16 krpm	42 krpm
d(O ₂)	125 sccm	250 sccm
Pola	0 V	500 V
d(TMA)	2 sccm	5 sccm

Sur le diagramme d'Ishikawa (**Figure C.2**) sont positionnés les facteurs principaux (en noir) et les facteurs fixés (en gris). Les **réponses** mesurées sont indiquées au bout de la flèche.

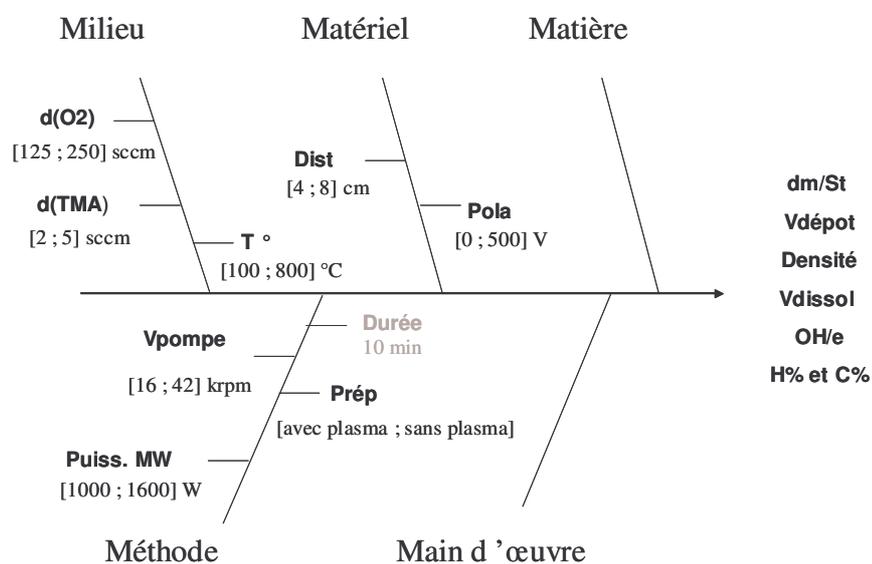


Figure C.2 Diagramme d'Ishikawa – Criblage des facteurs

I.2.2. MODELE EMPIRIQUE

Les objectifs d'un plan d'expériences renvoient à une forme spécifique de modélisation. Les modèles additifs sans couplage sont associés aux études de classement et de comparaison des effets moyens des facteurs (voir **Partie A**).

I.2.2.a. Forme générale du modèle additif sans couplage ⁽¹⁾

Un modèle additif sans couplage s'écrit sous la forme générale :

$$Y = C^{te} + \sum_{i=1}^k W_i$$

Dans le cas présent, en utilisant les abréviations des facteurs, on obtient :

$$Y = C^{te} + W_{\text{prep}} + W_{\text{temp}} + W_{\text{dist}} + W_{\text{puissMW}} + W_{\text{Vpompe}} + W_{\text{d(TMA)}} + W_{\text{pola}} + W_{\text{d(O2)}}$$

Cette écriture reflète bien les objectifs de l'étude, en identifiant le poids de chacun des niveaux des facteurs par le symbole W_i .

I.2.2.b. Nombre d'inconnues associées au modèle additif sans couplage

Une formule, permettant d'établir de façon simple et rapide le nombre p d'inconnues, est associée

au modèle additif sans couplage : $p = 1 + \sum_{i=1}^k (m_i - 1)$

La définition du nombre de facteurs ($k = 8$) et du nombre de modalités ($m_i = 2$) attribuées à chacun des facteurs, permet d'obtenir immédiatement :

p = 9 inconnus

Les plans d'expériences destinés à l'étude des effets des facteurs à deux modalités sont les plus répandus dans la littérature car ils sont les plus économiques, mais il ne faut pas en faire une généralisation. Chacune des décisions sur le choix d'une matrice doit se faire en considérant au mieux le problème posé. Prendre deux modalités alors que nous avons le choix de 3, 4 ou plus peut entraîner une perte d'information.

⁽¹⁾ « Association Expérimentique », I.U.T. d'Orléans Département OGP, BP 6729, 45067 Orléans Cedex 2

I.2.3. CONSTRUCTION DE LA MATRICE DU MODELE

I.2.3.a. Présentation de la matrice d'expérience

- La matrice d'expériences est construite comme suit :
- chaque expérience est sur une ligne d'où n lignes
- chaque facteur du plan est dans une colonne d'où k colonnes.

Afin de déterminer la constante C du modèle, une colonne uniquement composée de 1 est rajoutée à la matrice d'expériences. La matrice du modèle est donc un tableau de type $\mathbf{X} (n, k+1)$.

I.2.3.b. Objectif de la construction d'un plan de criblage

La construction d'un plan de criblage consiste à extraire du domaine expérimental un nombre suffisant N de combinaisons particulières afin d'estimer, de la meilleure façon possible, les p inconnues du modèle additif, tout en respectant au mieux les objectifs et les contraintes techniques et/ou économiques de l'étude.

Une des premières conditions nécessaires est de disposer d'un nombre N de traitements distincts supérieur ou égal au nombre p d'inconnues :

$$p \leq N \leq N_{\text{total}}$$

Comme nous l'avons dit dans la **Partie A** de ce manuscrit, il faut que les N traitements distincts retenus dans le dispositif étudié, permettent d'estimer les **effets moyens** de la meilleure façon possible (notion **d'orthogonalité**). Pour cela, on cherche à obtenir une incertitude à la fois la plus faible et la plus homogène possible pour l'estimation des inconnues du modèle, ce qui se traduit en quelque sorte, par une répartition la plus uniforme possible des essais à réaliser au sein du domaine étudié.

Les **plans de criblage** offrent uniquement des possibilités d'estimation des **effets moyens** des facteurs par l'intermédiaire d'un **modèle additif sans couplage**. Lorsque le nombre de modalités m_i est égal à **2** pour tous les facteurs, le nombre N de traitements expérimentaux distincts à réaliser est égal au **multiple de 4** immédiatement supérieur ou égal au nombre p d'inconnues à estimer. **R.L. Plackett et J.P. Burmann [3]** ont généralisé la méthode de construction de tels plans d'expériences. Repris et remodelés lors de la diffusion de la méthode **Taguchi [4]** liée à l'ingénierie robuste, les plans de **Plackett et Burmann** ont connu et rencontrent encore aujourd'hui un grand succès dans le monde industriel où, pour des raisons économiques, le nombre N de traitements expérimentaux distincts reste généralement limité (en général, il faut que N reste inférieur à 25).

Nous allons à présent détailler les différentes étapes de construction de ces plans d'expériences.

I.2.3.c. Principe de construction d'un plan de Plackett et Burmann⁽¹⁾

La construction d'un plan de **Plackett et Burmann** [3] est basée sur la duplication de lignes ou de colonnes contenant une alternance particulière de signes négatifs et positifs, par simple permutation circulaire. Cette construction s'effectue en **3 étapes** que nous allons expliquer ci-dessous.

Etape 1

Tout d'abord il faut **repérer la ligne génératrice** (figurant dans le **Tableau C.2**) pour laquelle le nombre **N** de traitements à réaliser, est immédiatement supérieur ou égal au nombre **p** d'inconnues à estimer. Dans le cas présent, nous retiendrons la ligne correspondant à **N = 12** puisque nous avons **p** égal à 9. Le nombre **N** est parfois appelé nombre d'**Hadamard**. Ces matrices sont également parfois appelées des matrices d'**Hadamard**⁽²⁾.

Tableau C.2 Lignes génératrices d'un plan de Plackett et Burmann

N=4	+	+	-																				
N=8	+	+	+	-	+	-	-																
N=12	+	+	-	+	+	+	-	-	-	+	-												
N=16	+	+	+	+	-	+	-	+	+	-	-	+	-	-	-								
N=20	+	+	-	-	+	+	+	+	-	+	-	+	-	-	-	-	+	+	-				
N=24	+	+	+	+	+	-	+	-	+	+	-	-	+	+	-	-	+	-	+	-	-	-	-

Etape 2

A présent, il faut transposer cette ligne dans la première ligne ou colonne d'une matrice de rang **N-1**, puis recopier cette première colonne par permutation circulaire. Il existe quatre permutations possibles : permutation par la gauche, permutation par la droite, permutation par le bas ou permutation par le haut. Nous allons expliquer la permutation par la droite.

La ligne génératrice ci-dessous sera donc la première ligne de notre tableau.

1 ^{ère} ligne	+	+	-	+	+	+	-	-	-	+	-
2 ^{ème} ligne	-	← +	+	-	+	+	+	-	-	-	+

Ensuite, il faut permuter le dernier signe de la ligne et le reporter sur la dernière colonne de la ligne suivante.

⁽¹⁾ « Association Expérimentique », I.U.T. d'Orléans département OGP, BP 6729, 45067 Orléans Cedex 2

⁽²⁾ Note de stage « Criblage et étude des facteurs », LPRAI B.P. n°7 13311 MARSEILLE Cedex 14

La ligne 1 comportant la liste de signes correspondant à la ligne génératrice ($N = 12$) est reportée dans la première ligne du **Tableau C.3**. Les flèches matérialisent la permutation circulaire utilisée pour générer l'ensemble des colonnes du tableau. Les colonnes sont repérées par des lettres allant de A à K et symbolisent 11 facteurs.

Etape 3

Enfin, il faut compléter la matrice (**Tableau C.3**) par une ligne exclusivement remplie de signes négatifs (**Tableau C.4**). Cette ligne (ligne 12) représente, dans cet exemple, le réglage de référence. Dans notre étude, il n'y a que 8 facteurs et pas 11 (nous avons donc enlevé les 3 dernières colonnes du **Tableau C.4**). L'arrangement orthogonal permet d'étudier les effets moyens des 8 facteurs ($k = 8$) à 2 modalités ($m_i = 2$) à partir de 12 traitements distincts ($N = 12$) au lieu des 256 expériences initiales.

Tableau C.3 Mode de construction par permutation circulaire droite

Lignes	A	B	C	D	E	F	G	H	I	J	K
1	+	+	-	+	+	+	-	-	-	+	-
	↘	↘	↘	↘	↘	↘	↘	↘	↘	↘	↘
2	-	+	+	-	+	+	+	-	-	-	+
3	+	-	+	+	-	+	+	+	-	-	-
4	-	+	-	+	+	-	+	+	+	-	-
5	-	-	+	-	+	+	-	+	+	+	-
6	-	-	-	+	-	+	+	-	+	+	+
7	+	-	-	-	+	-	+	+	-	+	+
8	+	+	-	-	-	+	-	+	+	-	+
9	+	+	+	-	-	-	+	-	+	+	-
10	-	+	+	+	-	-	-	+	-	+	+
11	+	-	+	+	+	-	-	-	+	-	+

Tableau C.4 Matrice d'expériences pour 11 facteurs à 2 niveaux

Lignes	A	B	C	D	E	F	G	H	I	J	K
1	+	+	-	+	+	+	-	-	-	+	-
2	-	+	+	-	+	+	+	-	-	-	+
3	+	-	+	+	-	+	+	+	-	-	-
4	-	+	-	+	+	-	+	+	+	-	-
5	-	-	+	-	+	+	-	+	+	+	-
6	-	-	-	+	-	+	+	-	+	+	+
7	+	-	-	-	+	-	+	+	-	+	+
8	+	+	-	-	-	+	-	+	+	-	+
9	+	+	+	-	-	-	+	-	+	+	-
10	-	+	+	+	-	-	-	+	-	+	+
11	+	-	+	+	+	-	-	-	+	-	+
12	-	-	-	-	-	-	-	-	-	-	-

Tableau C.5 Matrice d'expériences de l'étude

Lignes	Prép	T	Dist	Puiss. MW	Vpompe	d(TMA)	Pola	d(O2)
1	+	+	-	+	+	+	-	-
2	-	+	+	-	+	+	+	-
3	+	-	+	+	-	+	+	+
4	-	+	-	+	+	-	+	+
5	-	-	+	-	+	+	-	+
6	-	-	-	+	-	+	+	-
7	+	-	-	-	+	-	+	+
8	+	+	-	-	-	+	-	+
9	+	+	+	-	-	-	+	-
10	-	+	+	+	-	-	-	+
11	+	-	+	+	+	-	-	-
12	-	-	-	-	-	-	-	-

Nous noterons que la matrice **X** du modèle est de la forme :

$$X = \begin{pmatrix} + & + & + & - & + & + & + & - & - \\ + & - & + & + & - & + & + & + & - \\ + & + & - & + & + & - & + & + & + \\ + & - & + & - & + & + & - & + & + \\ + & - & - & + & - & + & + & - & + \\ + & - & - & - & + & - & + & + & - \\ + & + & - & - & - & + & - & + & + \\ + & + & + & - & - & - & + & - & + \\ + & + & + & + & - & - & - & + & - \\ + & - & + & + & + & - & - & - & + \\ + & + & - & + & + & + & - & - & - \\ + & - & - & - & - & - & - & - & - \end{pmatrix}$$

La matrice d'expériences obtenue après le rajout de la dernière ligne (**Tableau C.5**) possède des propriétés d'orthogonalité donc, elle garantit une estimation des effets moyens avec une incertitude minimale. L'estimation des effets moyens des facteurs s'effectue en comparant la moyenne des réponses observées pour chacune des modalités des facteurs. Cette comparaison est d'autant plus équitable que l'on introduit le même nombre d'observations dans chacune des moyennes et que ces observations sont nombreuses. C'est pour cette raison que l'on retrouve le même nombre de signes positifs et négatifs dans chaque colonne du tableau précédent.

Par ailleurs, chaque fois qu'un facteur est affecté d'un signe positif ou négatif dans une colonne, l'ensemble des autres facteurs apparaît un même nombre de fois avec un signe positif et avec un signe négatif dans les autres colonnes. Donc l'étude d'aucune modalité particulière n'est privilégiée.

I.2.4. EXPERIMENTATIONS

La mise en œuvre des expériences ne pose pas ici de problèmes particuliers ; l'affectation des variables de prédiction dans les colonnes du tableau représentant le plan d'expériences reprend l'ordre de présentation de ces variables. Les traitements expérimentaux sont réalisés suivant l'ordre des lignes du **Tableau C.6**. A cette étape, il est important de préparer le bon déroulement du plan d'expériences. Il est en particulier intéressant de retranscrire chacune des lignes du plan d'expérimentation sur une **fiche d'essai** (voir **Partie A**). Cette fiche facilite la préparation des traitements expérimentaux et limite les risques d'erreurs de lecture dans le tableau. Les traitements expérimentaux délicats à mettre en œuvre, doivent être réalisés en priorité au même titre que les réglages supposés occasionner les plus grandes variations des réponses observées.

Il est important de définir les traitements expérimentaux à réaliser avec le **groupe de travail** afin de les faire valider par les experts. De plus, il n'a pas été prévu ici de réaliser des répétitions, compte tenu du savoir-faire et de l'expérience des opérateurs.

Tableau C.6 Plan d'expérimentation

Expérience	Prép.	Temp.	Dist.	Puiss. MW	Vpompe	d(TMA)	Pola.	d(O2)
		°C	cm	W	krpm	sccm	V	sccm
1	sans	800	4	1600	42	5	0	125
2	plasma	800	8	1000	42	5	500	125
3	sans	100	8	1600	16	5	500	250
4	plasma	800	4	1600	42	2	500	250
5	plasma	100	8	1000	42	5	0	250
6	plasma	100	4	1600	16	5	500	125
7	sans	100	4	1000	42	2	500	250
8	sans	800	4	1000	16	5	0	250
9	sans	800	8	1000	16	2	500	125
10	plasma	800	8	1600	16	2	0	250
11	sans	100	8	1600	42	2	0	125
12	plasma	100	4	1000	16	2	0	125

I.2.5. ANALYSE GLOBALE DES RESULTATS D'ESSAI

Les résultats d'essais pour chacune des réponses sont reportés dans le **Tableau C.7**.

Pour chacune des réponses observées, il faut tout d'abord vérifier avec les expérimentateurs que leurs variations sont significatives. Nous pouvons observer rapidement l'étendue des réponses à l'aide d'un Box Plot (voir **Figure C.3**).

Tableau C.7 Résultats d'essais

Numéro d'essai	dm/St	Vdépôt	Densité	Vdissol	OH/e	C%	H%
	$\mu\text{g.cm}^{-2}\text{.min}^{-1}$	nm.min^{-1}		nm.s^{-1}	<i>u.a.</i>	%	%
1	81.5	284	3.69	495	74.6	38.9	21.4
2	29.2	145	4.13	4	78.5	49.5	28.3
3	14.4	68	2.71	48	79.9	50.4	30.8
4	11.8	115	1.41	4	1	35.1	21.6
5	29.4	195	2.01	36	126.6	32.7	17.6
6	30.9	150	2.77	71	169.9	63	44.2
7	14	54	3.45	123	24.7	66	38.5
8	31	146	2.71	171	92.9	60.6	31.1
9	2.9	15	2.4	11	77.1	32.9	18.4
10	2.4	33	1.5	33	104.9	34.8	18.9
11	18.3	79	3.23	41	149	26.1	16.5
12	14.5	119	1.7	317	213.4	40.7	26.8

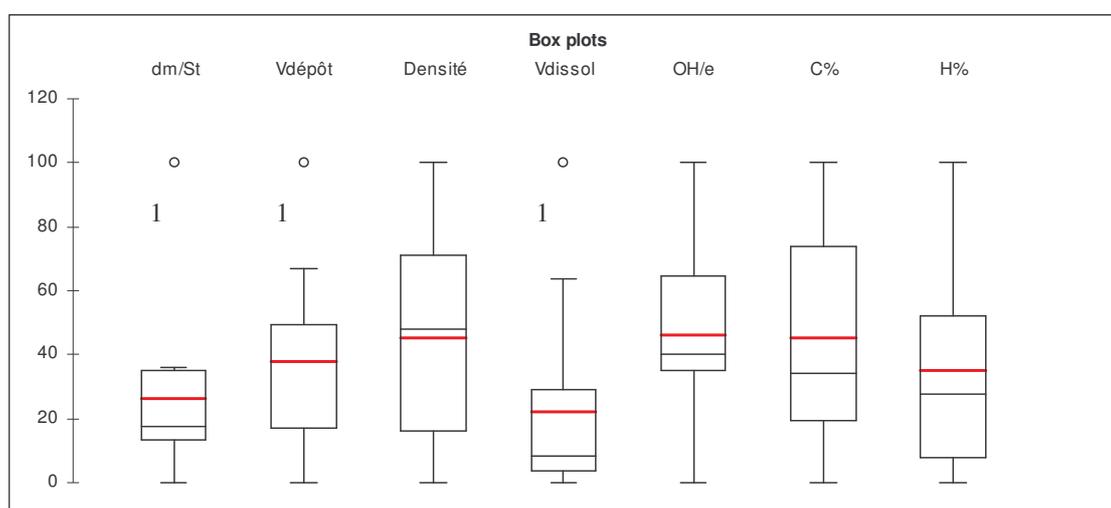


Figure C.3 Box Plot des réponses du criblage

A partir du Box plot (voir **Partie B**), nous pouvons déjà faire quelques observations :

- L'expérience 1 du plan correspond à une valeur atypique de la réponse dm/St. De plus, la médiane (trait noir) et la moyenne (en rouge) sont bien séparées ce qui traduit que l'étendue des résultats d'essais de la réponse dm/St n'est pas homogène : il y a plus d'expériences dont la valeur est inférieure à la moyenne des valeurs prises par dm/St,
- On observe également une valeur atypique (essai numéro 1) de la vitesse de dépôt. Les autres valeurs de cette réponse sont réparties de manière homogène dans le domaine expérimental (médiane confondue avec la moyenne).

- L'expérience 1 possède aussi une valeur atypique de la réponse V_{dissol} .
- Nous pouvons donc conclure que l'expérience 1 est une expérience atypique qui provoque des valeurs très élevées de trois réponses. Par ailleurs, nous constatons que les autres réponses (Densité, OH/e, C% et H%) ont des étendues homogène sans valeur aberrante.

Il convient maintenant de savoir si ces variations traduisent la manifestation **des effets des facteurs** au cours du plan d'expériences ou de la variabilité naturelle des réponses. L'analyse mathématique puis l'analyse statistique des résultats d'essais, vont permettre de répondre également à cette question.

I.2.6. ANALYSE MATHÉMATIQUE DES RESULTATS D'ESSAI

Dans le cas des plans de criblage, l'analyse mathématique des résultats consiste seulement à estimer les p inconnus du modèle. Elle fait appel le plus fréquemment à deux concepts :

- le **concept d'effet moyen** qui renvoie à l'utilisation d'une **grille de dépouillement**, particulièrement adaptée à l'analyse des résultats d'essais suite à un plan d'expériences respectant un arrangement orthogonal⁽¹⁾,
- le **concept de coefficient** qui ne se limite pas aux arrangements orthogonaux. Il se base sur l'emploi d'un modèle polynomial adapté pouvant faciliter par la suite la mise en œuvre de certains tests statistiques.
- En présence de k facteurs à deux modalités ($m_i = 2$) et grâce à l'utilisation d'un arrangement orthogonal, l'estimation des p inconnues du modèle à partir de ces deux concepts, conduit à la même interprétation des résultats. Nous allons cependant préciser les différents concepts.

I.2.6.a. Concept d'effet moyen⁽¹⁾

L'**effet moyen** d'un facteur est défini par la variation observée ou modélisée de la réponse caractérisant le phénomène à étudier, entre deux états distincts correspondant aux modalités du facteur. L'utilisation d'une grille de dépouillement facilite la restitution de l'information recherchée sous la forme graphique du tracé des effets.

Mais, en présence d'un grand nombre de facteurs, la construction d'une grille de dépouillement et son utilisation peuvent comporter des risques d'erreur, donc nous ne l'utiliserons pas ici.

Avec ce concept, il est possible de calculer autant de combinaisons linéaires indépendantes L_i qu'il y a de colonnes dans la matrice des effets, chacune de ces combinaisons linéaires représentant un effet b_i .

L'estimation de ces effets s'obtient en multipliant les éléments de la colonne X_i par les éléments correspondants du vecteur Y étudié, puis en faisant la somme algébrique et en divisant par le nombre d'expériences N .

Pour exemple, nous allons prendre la réponse $V_{\text{dépôt}}$ que nous allons analyser (voir **Tableau C.8**).

A partir des douze résultats expérimentaux, nous pouvons calculer douze informations correspondant à douze combinaisons linéaires. Les huit premières combinaisons vont nous permettre de calculer une estimation des effets des facteurs (b_i).

$Y = \begin{cases} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ y_{10} \\ y_{11} \\ y_{12} \end{cases}$

Tableau C.8 Réponse $V_{\text{dépôt}}$

Vdépôt	Y	N°Exp
284	y1	1
145	y2	2
68	y3	3
115	y4	4
195	y5	5
150	y6	6
54	y7	7
146	y8	8
15	y9	9
33	y10	10
79	y11	11
119	y12	12

La première combinaison linéaire, L_0 , est calculée comme l'indique cette équation :

$$L_0 = (+ y_1 + y_2 + y_3 + y_4 + y_5 + y_6 + y_7 + y_8 + y_9 + y_{10} + y_{11} + y_{12})/12 = b_0$$

Il faut noter que la **constante b_0** du modèle correspond à la **moyenne arithmétique des résultats d'essai**.

Les lignes suivantes sont calculées comme l'indique le fichier **Excel** de la **Figure C.4**.

(1) « Association Expérimentique », I.U.T. d'Orléans département OGP, BP 6729, 45067 Orléans Cedex 2

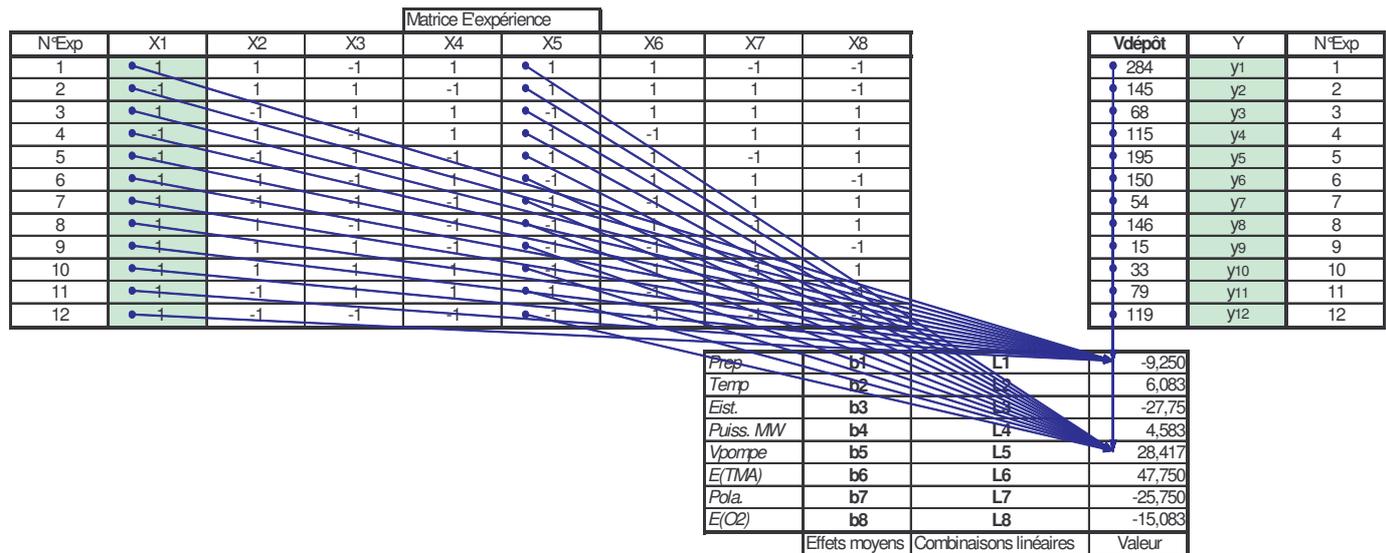


Figure C.4 Calcul des effets moyens (EXCEL)

Ce qui donne par exemple pour la combinaison L_1 :

$$L_1 = (+ y_1 - y_2 + y_3 - y_4 - y_5 - y_6 + y_7 + y_8 + y_9 - y_{10} + y_{11} - y_{12})/12 = b_1$$

Et ainsi de suite pour $L_2, L_3, L_4, L_5, L_6, L_7$ et L_8 qui représentent respectivement les effets $b_2, b_3, b_4, b_5, b_6, b_7$ et b_8 . Après calcul, on obtient les résultats reportés dans le **Tableau C.9**.

Tableau C.9 Calcul des effets moyens pour la réponse $V_{dépôt}$

	Effets moyens	Combinaisons linéaires	Valeur
Prep	b1	L1	-9,250
Temp	b2	L2	6,083
Eist.	b3	L3	-27,75
Puiss. MW	b4	L4	4,583
Vpompe	b5	L5	28,417
E(TMA)	b6	L6	47,750
Pola.	b7	L7	-25,750
E(O2)	b8	L8	-15,083

I.2.6.b. Concept de coefficient - Modèle polynomial

Cette approche repose sur l'utilisation d'un polynôme du premier degré de telle sorte que, pour l'expression du résultat d'essai du $i^{ème}$ traitement repéré par y_i , on puisse écrire :

$$\hat{y}_i = b_0 + \sum_{j=1}^k b_j x_{i,j} \tag{1}$$

En remplaçant les indices par les abréviations des facteurs de l'étude, nous obtenons :

$$y_i = b_0 + b_{prep} X_{i,prep} + b_{temp} X_{i,temp} + b_{dist} X_{i,dist} + b_{puissMW} X_{i,puissMW} + b_{Vpompe} X_{i,Vpompe} + b_{d(TMA)} X_{i,d(TMA)} + b_{pola} X_{i,pola} + b_{d(O2)} X_{i,d(O2)} + e_i$$

A titre d'exemple, le résultat d'essai du 1^{er} traitement du plan d'expériences peut s'écrire :

$$\hat{y}_1 = b_0 + b_{\text{prep}} + b_{\text{temp}} - b_{\text{dist}} + b_{\text{puissMW}} + b_{\text{Vpompe}} + b_{\text{d(TMA)}} - b_{\text{pola}} - b_{\text{d(O2)}}$$

En utilisant les valeurs de x_{ij} proposées par la matrice d'expériences et la définition du modèle **(1)**, on peut écrire le système suivant :

$$\begin{aligned} \hat{y}_1 &= b_0 + b_{\text{prep}} + b_{\text{temp}} - b_{\text{dist}} + b_{\text{puissMW}} + b_{\text{Vpompe}} + b_{\text{d(TMA)}} - b_{\text{pola}} - b_{\text{d(O2)}} \\ \hat{y}_2 &= b_0 - b_{\text{prep}} + b_{\text{temp}} + b_{\text{dist}} - b_{\text{puissMW}} + b_{\text{Vpompe}} + b_{\text{d(TMA)}} + b_{\text{pola}} - b_{\text{d(O2)}} \\ \hat{y}_3 &= b_0 + b_{\text{prep}} - b_{\text{temp}} + b_{\text{dist}} + b_{\text{puissMW}} - b_{\text{Vpompe}} + b_{\text{d(TMA)}} + b_{\text{pola}} + b_{\text{d(O2)}} \\ \hat{y}_4 &= b_0 - b_{\text{prep}} + b_{\text{temp}} - b_{\text{dist}} + b_{\text{puissMW}} + b_{\text{Vpompe}} - b_{\text{d(TMA)}} + b_{\text{pola}} + b_{\text{d(O2)}} \\ \hat{y}_5 &= b_0 - b_{\text{prep}} - b_{\text{temp}} + b_{\text{dist}} - b_{\text{puissMW}} + b_{\text{Vpompe}} + b_{\text{d(TMA)}} - b_{\text{pola}} + b_{\text{d(O2)}} \\ \hat{y}_6 &= b_0 - b_{\text{prep}} - b_{\text{temp}} - b_{\text{dist}} + b_{\text{puissMW}} - b_{\text{Vpompe}} + b_{\text{d(TMA)}} + b_{\text{pola}} - b_{\text{d(O2)}} \\ \hat{y}_7 &= b_0 + b_{\text{prep}} - b_{\text{temp}} - b_{\text{dist}} - b_{\text{puissMW}} + b_{\text{Vpompe}} - b_{\text{d(TMA)}} + b_{\text{pola}} + b_{\text{d(O2)}} \\ \hat{y}_8 &= b_0 + b_{\text{prep}} + b_{\text{temp}} - b_{\text{dist}} - b_{\text{puissMW}} - b_{\text{Vpompe}} + b_{\text{d(TMA)}} - b_{\text{pola}} + b_{\text{d(O2)}} \\ \hat{y}_9 &= b_0 + b_{\text{prep}} + b_{\text{temp}} + b_{\text{dist}} - b_{\text{puissMW}} - b_{\text{Vpompe}} - b_{\text{d(TMA)}} + b_{\text{pola}} - b_{\text{d(O2)}} \\ \hat{y}_{10} &= b_0 - b_{\text{prep}} + b_{\text{temp}} + b_{\text{dist}} + b_{\text{puissMW}} - b_{\text{Vpompe}} - b_{\text{d(TMA)}} - b_{\text{pola}} + b_{\text{d(O2)}} \\ \hat{y}_{11} &= b_0 + b_{\text{prep}} - b_{\text{temp}} + b_{\text{dist}} + b_{\text{puissMW}} + b_{\text{Vpompe}} - b_{\text{d(TMA)}} - b_{\text{pola}} - b_{\text{d(O2)}} \\ \hat{y}_{12} &= b_0 - b_{\text{prep}} - b_{\text{temp}} - b_{\text{dist}} - b_{\text{puissMW}} - b_{\text{Vpompe}} - b_{\text{d(TMA)}} - b_{\text{pola}} - b_{\text{d(O2)}} \end{aligned}$$

Il est courant de retranscrire ce système sous forme matricielle :

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ y_{10} \\ y_{11} \\ y_{12} \end{pmatrix} = \begin{pmatrix} + & + & + & - & + & + & + & - & - \\ + & - & + & + & - & + & + & + & - \\ + & + & - & + & + & - & + & + & + \\ + & - & + & - & + & + & - & + & + \\ + & - & - & + & - & + & + & - & + \\ + & - & - & - & + & - & + & + & - \\ + & + & - & - & - & + & - & + & + \\ + & + & + & - & - & - & + & - & + \\ + & + & + & + & - & - & - & + & - \\ + & - & + & + & + & - & - & - & + \\ + & + & - & + & + & + & - & - & - \\ + & - & - & - & - & - & - & - & - \end{pmatrix} \begin{pmatrix} b_0 \\ b_{\text{prep}} \\ b_{\text{temp}} \\ b_{\text{dist}} \\ b_{\text{puissMW}} \\ b_{\text{Vpompe}} \\ b_{\text{d(TMA)}} \\ b_{\text{pola}} \\ b_{\text{d(O2)}} \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \\ e_9 \end{pmatrix}$$

Soit encore sous une forme plus générale (aux résidus près) :

$$(Y) = (X)(B) \quad (\text{Voir } \mathbf{Partie A}).$$

Dans le cas où cette matrice est carrée de rang N , on l'appelle matrice d'Hadamard. Elle doit vérifier la relation suivante : $({}^tX)(X) = N(I_N)$

Dans cette dernière expression, les matrices $({}^tX)$ et (I_N) représentent respectivement la transposée de la matrice du modèle (X) et la matrice identité de rang N .

En désignant par $(X)^{-1}$ l'inverse de la matrice du modèle (X) , on peut écrire dans le cas présent :

$$(X)^{-1} = \frac{1}{N}({}^tX) \quad (2)$$

Il est alors possible de calculer les coefficients grâce à l'équation **(3)** ci dessous :

$$(B) = (X)^{-1}(Y) \text{ soit d'après l'équation (2) donc } (B) = \frac{1}{N}({}^tX)(Y) \quad (3)$$

Dans notre étude, le plan d'expériences n'est pas **saturé** : le nombre p d'inconnues à estimer n'est pas égal au nombre N de traitements distincts réalisés (on rappelle que $p = 8$ et $N = 12$ donc $p \neq N$). La matrice (X) n'est donc plus une matrice carrée, il faut alors recourir à l'écriture matricielle de la **méthode des moindres carrés**.

L'équation **(4)** permet ainsi de trouver les coefficients avec la méthode des moindres carrés :

$$(B) = ({}^tXX)^{-1}({}^tX)(Y) \quad (4)$$

En utilisant la relation **(2)**, on obtient à nouveau : $(B) = \frac{1}{N}({}^tX)(Y)$.

L'estimation des coefficients du modèle obéit à la même équation mathématique, que le plan d'expériences soit saturé ou non. A un coefficient multiplicatif près, cette équation traduit l'expression du produit scalaire entre les colonnes de la matrice du modèle (X) et le vecteur des résultats d'essai (Y) . La mise en œuvre de ce calcul est souvent accessible à partir d'un simple tableur informatique.

Tableau C.10 Estimation des coefficients b_i

		Valeur
Moyenne	b0	116.9
Prep	b1	-9.3
Temp	b2	6.1
Dist.	b3	-27.8
Puiss. MW	b4	4.6
Vpompe	b5	28.4
d(TMA)	b6	47.8
Pola.	b7	-25.8
d(O2)	b8	-15.1

Les estimateurs b_i obtenus par le concept des **coefficients** (calcul du logiciel **NemrodW**) sont reportés dans le **Tableau C.10**. Les valeurs sont proches de celles obtenues à partir du concept des effets moyens.

I.2.7. ANALYSE GRAPHIQUE DU MODELE

Une communication des résultats sous forme graphique favorise les échanges entre les différents participants d'une réunion de travail et facilite souvent l'interprétation des informations obtenues.

Il existe différents types de graphiques pour illustrer les effets moyens des facteurs : les plus courants sont le diagramme en bâtons et le diagramme de Pareto.

I.2.7.a. Le diagramme en bâton- tracé des effets moyens

La construction du tracé des effets moyens utilise les valeurs des coefficients reportés dans le **Tableau C.10**, présentée sous forme de « diagramme en bâtons » ; la surface de chaque bâton est proportionnelle à la valeur de l'effet de ce facteur sur la variation de la réponse.

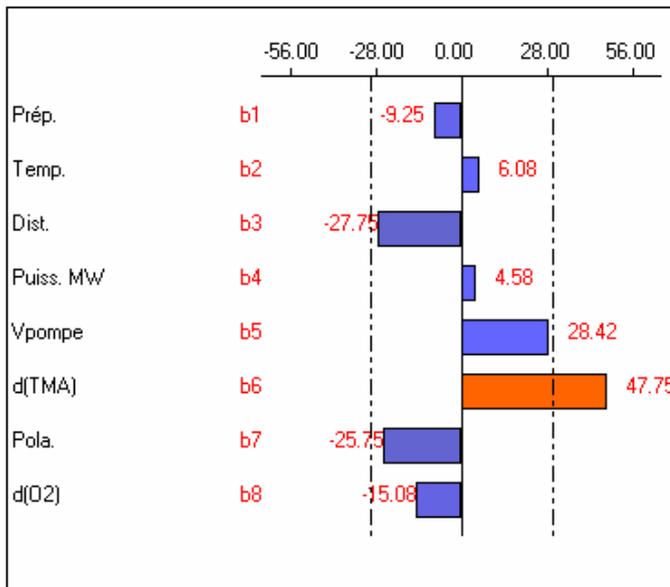


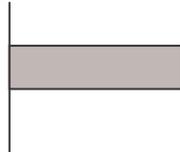
Figure C.5

Etude de la réponse $V_{\text{dépôt}}$

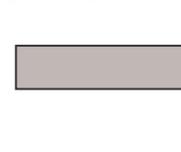
A savoir que :

Plus le bâton est grand, plus le poids du facteur est important.

Effet b_j positif : passage du - au +.



Effet b_j négatif : passage du + au -.



Limites de confiance

Sur ce diagramme, les **limites de confiance** indiquées par deux **droites en pointillés** correspondent à un seuil de signification.

Il existe des outils d'aide à l'interprétation des résultats qui permettent de fixer ces limites dites de « **signification** ». Nous verrons certains de ces outils dans l'analyse statistique des résultats.

Interprétation

Pour la réponse traduisant la **vitesse de dépôt**, le tracé des effets moyens (**Figure C.5**) fait ressortir immédiatement les facteurs importants : la variable d(TMA) a le plus grand effet (positif) sur la variation de la vitesse de dépôt, de plus, la température, la puissance MW et la préparation apparaissent comme des facteurs ayant une action beaucoup moins importante. L'expérience des opérateurs et la connaissance du groupe de travail ont confirmé ces résultats.

I.2.7.b. Diagramme de PARETO

Il est possible de décomposer la variation d'une réponse à partir des **contributions** apportées par chacun des facteurs dans un modèle. L'écriture polynomiale (5) facilite la définition des

contributions d'un facteur j à partir de la relation suivante :
$$CTR_j = \frac{b_j^2}{\sum_{j=1}^k b_j^2} \quad (5).$$

Les contributions des facteurs sont alors ordonnées par **ordre croissant** puis représentées sous forme de diagramme en bâtons. Le diagramme de Pareto est complémentaire au tracé des effets moyens. Il permet de déterminer les facteurs influents par ordre de contribution décroissante. Le diagramme de Pareto relatif à la vitesse de dépôt ($V_{\text{dépôt}}$) (**Figure C.6**) renforce la prédominance du rôle du facteur d(TMA) qui apparaît le premier avec un bâton beaucoup plus grand que les autres. Il est également possible de tracer la fonction **cumulée** de ces contributions, les valeurs normées de la **Figure C.6** sont reprises et additionnées au fur et à mesure (voir **Figure C.7**).

Ainsi, nous pouvons voir qu'à eux seuls, les quatre premiers facteurs (**d(TMA)**, **Vpompe**, **Dist.et Pola**) expliquent **plus de 90%** de la variation de la réponse $V_{\text{dépôt}}$.

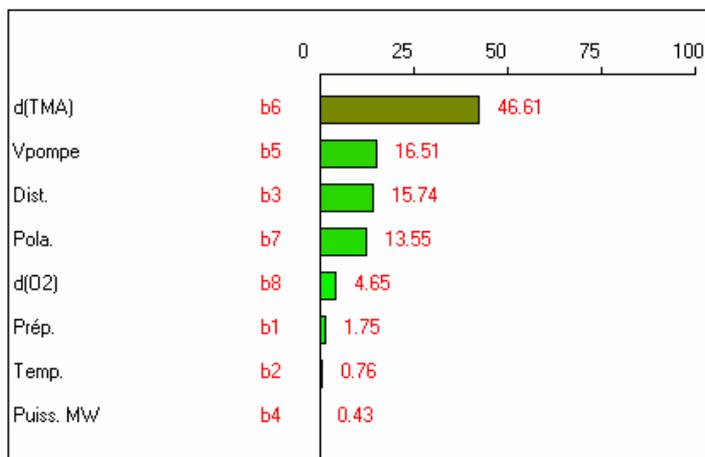


Figure C.6
Diagramme de Pareto pour la réponse $V_{\text{dépôt}}$

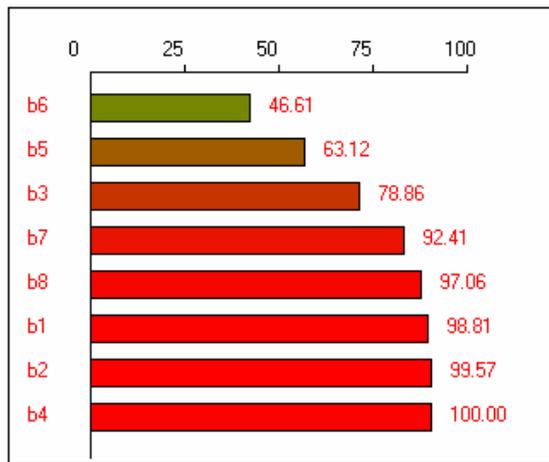


Figure C.7
Diagramme de Pareto cumulé pour la
réponse $V_{\text{dépôt}}$

I.2.8. ANALYSE STATISTIQUE DU MODELE

I.2.8.a. Objectif

L'analyse statistique représente une aide à l'interprétation des résultats. Dans le cas présent, l'analyse statistique permet d'identifier les effets probablement « **significatifs** » ou « **actifs** » des facteurs. L'analyse statistique fait appel à la mise en œuvre de **tests statistiques**, procédures bien définies qui permettent de traduire, à partir d'une probabilité, la pertinence d'un modèle et le caractère significatif des effets moyens qu'il permet d'identifier. Les procédures de construction des tests statistiques renvoient à des hypothèses et nécessitent la connaissance de la variabilité naturelle des résultats d'essai, rarement disponible dans une étude. En effet, le manque de temps et d'argent concernant une étude n'autorise pas souvent les possibilités de répétition d'expériences. Dans ces cas là, il est donc impossible de mettre en œuvre de façon efficace des tests statistiques et on s'en remet aux experts pour l'interprétation. Mais, ces constats ont suscité depuis longtemps le développement de nombreuses méthodes autorisant une approche d'analyse statistique en l'absence de répétition. En revanche, dès qu'il est possible, il est conseillé de faire des répétitions.

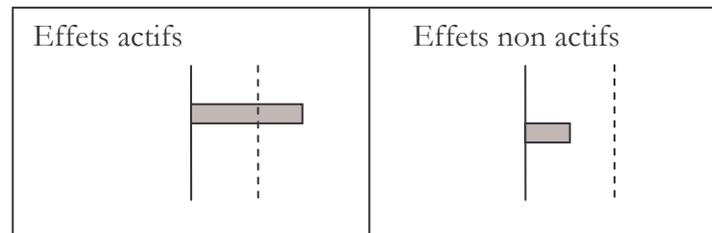
Nous présenterons dans ce chapitre deux méthodes :

- la méthode de **Lenth**,
- la méthode de la droite de **Henry**.

I.2.8.b. Méthode de Lenth

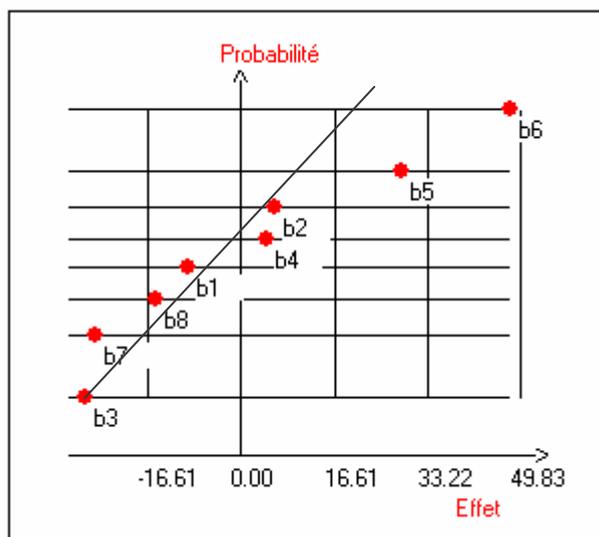
La méthode de **Lenth** est plus récente ; elle consiste à estimer une pseudo erreur-type pour mettre en œuvre un test statistique dont le résultat se traduit par le tracé des limites de signification sur le diagramme en bâtons (**Figure C.5**). L'application informatisée de cette méthode porte sur les estimateurs des coefficients.

Les coefficients b_j dont les valeurs estimées sont situées à l'extérieur des limites correspondent à des effets actifs. Les coefficients b_j dont les valeurs estimées sont situées à l'intérieur des limites correspondent à des effets non actifs. Entre ces deux limites, il est nécessaire de disposer d'informations complémentaires pour apporter une conclusion et se prononcer sur les effets moyens des facteurs.

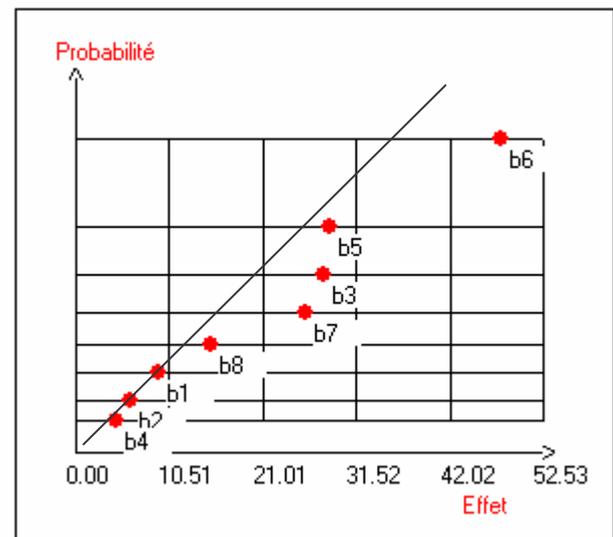


I.2.8.c. Méthode de la droite de HENRY

Si l'on représente la fonction de répartition des effets (b_j ou $|b_j|$) sur du papier Gausso-arithmétique, nous obtenons deux graphes : le graphe **Normal Plot** avec les b_j et le graphe **Half Normal Plot** avec les $|b_j|$ (voir **Figure C.8**).



Normal Plot



Half Normal Plot

Figure C.8 Droite de HENRY (Réponse $V_{\text{dépot}}$)

Les effets qui s'écartent de la droite sur le Normal plot (en dessous ou haut-dessus de la droite, selon que les effets sont positifs ou négatifs) sont considérés comme « **probablement actifs** ». Le graphe du Half Normal plot permet de mieux distinguer les coefficients actifs mais il n'en donne pas le sens puisque l'on considère ici la valeur absolue du coefficient.

Le coefficient b_6 est probablement actif ce qui est en accord avec les résultats obtenus avec la méthode de Lenth (Figure C.7).

I.2.9. BILAN ET CONCLUSION

Tous les résultats graphiques de cette étude sont contenus dans la thèse d'H. Hidalgo [1], nous ne retiendrons ici que les résultats nécessaires à la suite de l'étude et plus particulièrement les conclusions de la Figure C.9. Ce diagramme de Pareto est réalisé à partir des poids moyens des facteurs cumulés de l'ensemble des réponses. Il détermine ainsi les facteurs les plus importants (seul le coefficient C% a été pris en compte pour la sortie homogénéité).

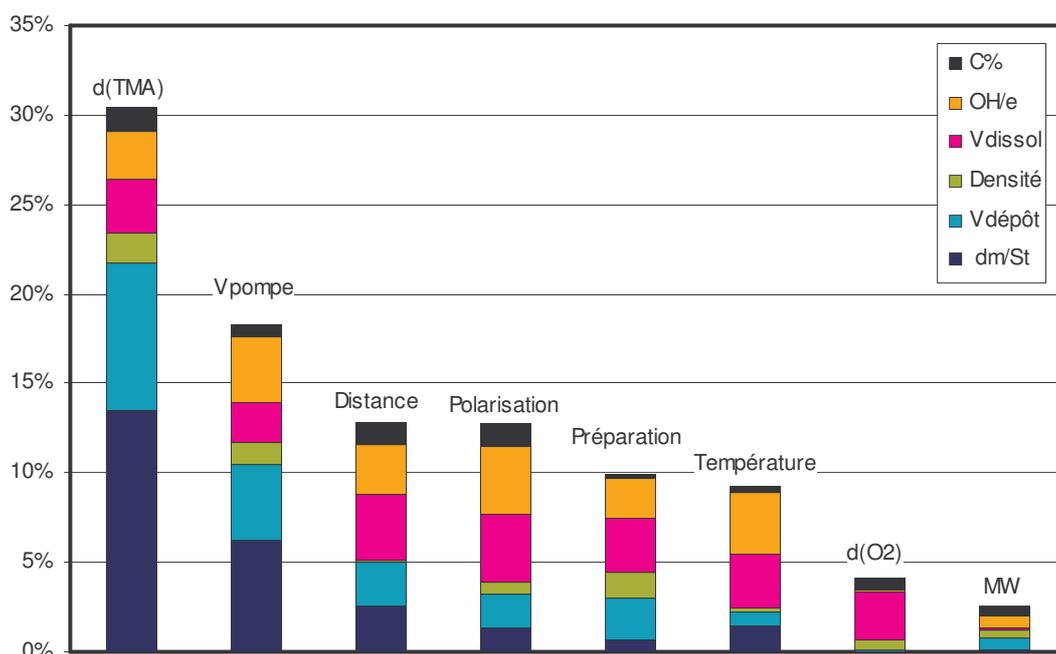


Figure C.9 Diagramme de Pareto (réponses cumulées)

La Figure C.9 permet de regrouper les facteurs en trois classes :

- les paramètres importants : $d(TMA)$ et $Vpompe$,
- les paramètres non négligeables : *Distance* (dist.), *Polarisation* (Pola.), *Préparation* (Prep.) et *Température* (Temp.),
- les paramètres négligeables : Puissance MW et Débit d'oxygène $d(O_2)$.

Les paramètres les plus importants dans le contrôle du dépôt et des propriétés physico-chimiques des couches sont donc : le débit de TMA , la vitesse de rotation de la pompe, la distance injecteur/substrat et la polarisation.

Compte tenu de l'effet très important du *débit de TMA* sur les vitesses de dépôt, les influences des autres facteurs pourraient être masquées. Ces raisons nous ont conduit à **fixer le débit de TMA** à la valeur (3 sccm) pour la suite de l'étude.

Le poids du facteur *préparation de surface* (plasma d'argon (50 sccm, 1 Pa, 1000 W, 500 V, 10 min)) n'est pas négligeable. En particulier, cette étape de pré-dépôt se traduit systématiquement par un effet positif sur les caractéristiques des couches. Nous avons donc choisi de réaliser systématiquement cette opération, comme d'ailleurs pour l'étude paramétrique.

La *puissance micro-onde* ainsi que le *débit d'oxygène* ayant peu d'influence, ils peuvent être fixés à une valeur quelconque dans la plage explorée : nous avons choisi de nous rapprocher des conditions expérimentales de l'étude paramétrique et nous avons fixé ces deux paramètres à des valeurs respectives de 1600 W et 125 sccm.

N'oublions pas un des paramètres qui a été fixé au départ, c'est la *durée du dépôt* (10 minutes).

Le **Tableau C.11** résume les valeurs prises pour les facteurs fixés pour la suite de l'étude.

Tableau C.11 Facteurs fixés pour le reste de l'étude

Débit O ₂	Débit TMA	Puissance MW	Préparation	Durée dépôt
125 sccm	3 sccm	1600 W	Avec plasma	10(min)

En conclusion, ce plan de criblage met bien en évidence les paramètres qui ont peu d'influence sur le système et permet de choisir au mieux, ceux qui semblent les plus intéressants pour une étude ultérieure plus précise. Un **plan d'expériences à surface de réponses** permettra d'obtenir un réglage plus affiné de ces facteurs correspondant à une optimisation des propriétés des films. C'est ce que nous allons d'écrire dans la suite de cette partie.

I.3. OPTIMISATION – PLANS A SURFACES DE REPONSES

I.3.1. INTRODUCTION

Il existe différentes méthodes d'optimisation parmi lesquelles on peut citer les plans d'expériences pour l'étude des **surfaces de réponses**. La plupart de nos études seront des problèmes d'optimisation. Cette méthode permet la recherche des niveaux optimaux des facteurs pour atteindre un niveau de réponse souhaité.

Les plans d'expériences pour l'étude des surfaces de réponses constituent souvent une deuxième étape, après un plan de criblage destiné à identifier les facteurs les plus influents.

I.3.2. FACTEURS ET NIVEAUX

Les 4 facteurs utilisés ont été déterminés par le plan de criblage : la **polarisation**, la **distance** Injecteur/substrat, la **température** et la **pression** (dissimulée derrière la *vitesse de rotation de la pompe* et le *débit d'oxygène*). Le facteur pression remplace donc le facteur V_{pompe} .

I.3.2.a. Contraintes pour la pression :

La **pression** est le facteur le plus difficile à régler. En effet, une modification de la vitesse de rotation de la pompe turbo moléculaire ne permet pas systématiquement d'obtenir la pression désirée. Par contre, une légère variation (maximum 4%) du débit d'oxygène (que l'on considérera malgré tout constant) permet d'ajuster plus finement la pression à la valeur désirée (Tableau C.12).

Tableau C.12 Valeur de la pression pour différents couples vitesse de la pompe /débit d'oxygène

Vitesse de rotation de la pompe (tr/min)	Débit d'oxygène (sccm)	Pression (Pa)
40000	125	1,0
22000	130	3,3
19000	125	5,5
17000	119	7,8
16000	125	10,0

A présent, on définit le domaine expérimental qui détermine les limites de variation des facteurs U_{min} et U_{max} à l'intérieur desquelles on définira des niveaux U_i (voir **Tableau C.13**).

Tableau C.13 Domaine expérimental

Notation	Facteur	Unité	Centre	Pas de variation	Umin	Umax
U1	Pression	Pa	5,5	4,5	1	10
U2	Polarisation	V	250	250	0	500
U3	T°	°C	450	350	100	800
U4	Distance	cm	6	2	4	8

I.3.3. REPONSES POUR LE PLAN D'OPTIMISATION

Les sorties correspondant aux vitesses de dépôt (dm/St , V_{depot}) et aux marqueurs de qualité du film ($densité$, V_{dissol} , OH/e) sont conservées. Mais, contrairement au plan précédent, l'étude sur l'homogénéité en épaisseur des couches ne sera pas réalisée.

Au lieu d'un seul substrat (\varnothing 10 cm), nous avons introduit dans le réacteur des échantillons de taille standard (2cm \times 2cm \times 500 μ m) en silicium en compagnie de substrats métalliques faisant l'objet d'une étude conjointe. Par contre, l'*adhérence* des films au substrat a été prise en compte.

En plus des caractéristiques des films, la composition de la phase gazeuse sera suivie par spectrométrie d'émission. Pour plus de précision sur ces réponses, il faut se référer à la thèse d'**H.Hidalgo** [1]. Dans cette étude nous allons étudier plus précisément la famille de réponses liées aux **propriétés du film**. Tous les facteurs fixés (en gris) ou principaux (en noir) ainsi que les réponses mesurées sont établies de manière plus claire sur le diagramme d' **Ishikawa** de la **Figure C.10**.

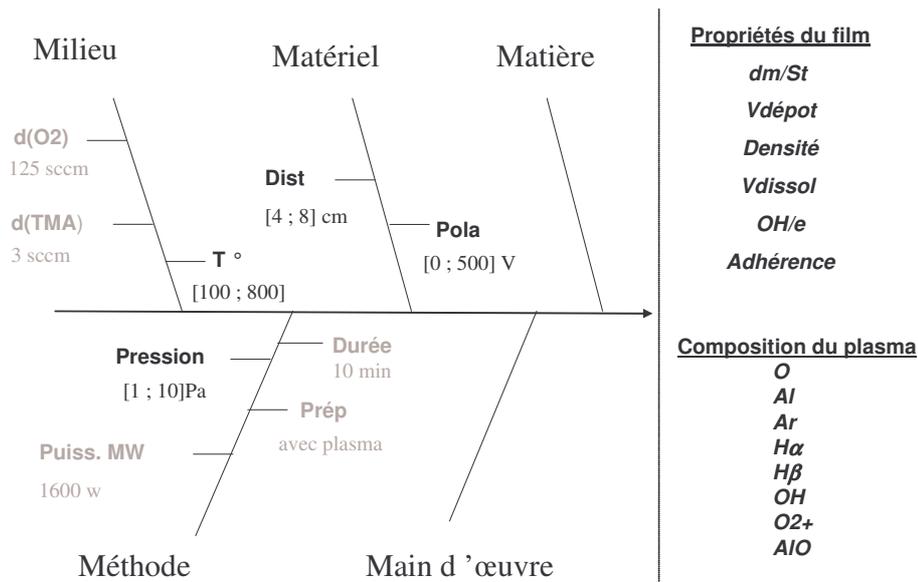


Figure C.10 Digramme d'ISHIKAWA – Optimisation

I.3.4. GEOMETRIE DU DOMAINE EXPERIMENTAL⁽¹⁾

I.3.4.a. Définition du domaine expérimental

La définition du niveau minimal et du niveau maximal pour chacun des facteurs (**Tableau C.13**) permet de délimiter le domaine expérimental. Dans ces plans d'expériences (plan à **surfaces de réponses**), les niveaux des facteurs sont définis, le plus souvent, à partir de critères algébriques ou géométriques lors de la construction de la matrice d'expériences.

I.3.4.b. Géométrie sphérique ou cubique

La matrice d'expériences peut être établie au sein d'un **domaine sphérique** ou **cubique** ; ce choix est essentiellement lié à la présence ou non d'un réglage de référence, définissant généralement le centre du domaine.

- *Domaine sphérique* : l'expérimentateur désire explorer la zone du plan autour de son centre défini comme point de référence.

- *Domaine cubique* : l'expérimentateur désire prévoir les variations de la variable de réponse pour des combinaisons des niveaux extrêmes des facteurs.

Nous avons choisi ici un domaine sphérique.

I.3.5. MODELE EMPIRIQUE

On rappelle que l'étude des surfaces de réponse est associée à l'utilisation des polynômes de degré 2 qui s'écrivent sous la forme générale suivante :

$$\hat{y}_i = b_0 + \sum_{i=1}^k b_i x_i + \sum_{i=1}^k b_{ii} x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^k b_{ij} x_i x_j$$

Dans le cas présent, le modèle polynomial de forme générale s'écrit donc :

$$\begin{aligned} y_i = & b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + b_{11} x_1^2 + b_{22} x_2^2 + b_{33} x_3^2 + b_{44} x_4^2 \\ & + b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{23} x_2 x_3 + b_{14} x_1 x_4 + b_{24} x_2 x_4 + b_{34} x_3 x_4 \\ & + \text{résidus} \end{aligned} \quad (5)$$

Le nombre d'inconnues d'un polynôme de degré 2 est établi à partir de la formule suivante :

$$p = \frac{(k+2)!}{k!2!} \Rightarrow p = \frac{(k+2)(k+1)}{2}$$

Soit, dans cet exemple : $p = \frac{(4+2)!}{4!2!} \Rightarrow p = 15$ inconnues

L'écriture du modèle permet de définir le nombre d'inconnues à estimer à partir de l'analyse des résultats d'essais, suite au résultat du plan d'expérimentation. L'analyse mathématique des résultats d'essais s'appuie sur la **méthode des moindres carrés** pour l'estimation des inconnues du modèle que nous avons déjà expliquée dans le paragraphe **I.2.6.b** et dans la **Partie A**.

I.3.6. CONSTRUCTION DU PLAN D'EXPERIENCES

Le nombre de lignes d'une matrice d'expériences destinée à l'étude des surfaces de réponse croît très rapidement en fonction du nombre de facteurs. Ainsi, l'emploi de cette méthode se limite à l'étude de quatre ou cinq facteurs continus tout au plus.

Le plan d'expériences illustré dans ce chapitre a été proposé par **D. H. Doehlert** en 1970, plus couramment appelé matrice de **Doehlert**.

I.3.6.a. Objectif de la construction d'une matrice de Doehlert

Par définition, comme pour tout plan d'expériences, on doit respecter la relation :

$$p \leq N \quad \text{dans notre cas} \quad N \geq 15$$

La géométrie **sphérique** ou **cubique** ainsi que le nombre de niveaux qu'il est possible de réaliser influence la construction de la matrice d'expériences. En effet, le nombre de niveaux influence la restitution des courbures de la réponse en fonction de leurs variations.

Nous allons préciser dans ce chapitre la méthode de construction des plans d'expériences proposés par **D. H. Doehlert** en 1970, à partir de considérations uniquement géométriques.

Les matrices de **Doehlert** sont des matrices asymétriques pour lesquelles tous les facteurs ne possèdent pas le même nombre de niveaux et dont le nombre de traitements expérimentaux distincts obéit à la relation suivante : $N = k^2 + k + 1$

$$N = 4^2 + 4 + 1 = 21 \text{ expériences à réaliser}$$

I.3.6.b. Principe de construction d'un réseau uniforme de Doehlert⁽¹⁾

De manière à illustrer ce principe, nous allons simplifier en commençant la construction d'un domaine défini par deux facteurs X_1 et X_2 . N'oublions pas qu'ici nous fonctionnons avec les variables codées des facteurs X_1 et X_2 (Voir **Partie A**).

Les essais sont disposés sur un cercle de rayon unitaire (voir **Figure C.11**), à l'exception du centre du domaine qui représente le réglage ou la formule de référence, à partir duquel on s'appuie pour construire un **simplexe** initial. Un **simplexe** est une figure comportant un sommet de plus que la dimension de l'espace dans lequel il est représenté. Ici, on est dans un espace à deux dimensions, donc le simplexe sera une figure à trois sommets.

Dans le cas des réseaux uniformes de **Doehlert**, on utilise un simplexe régulier, c'est-à-dire un triangle équilatéral, avec une longueur d'arrête égal à un. Le **simplexe** est un triangle de trois expériences construit selon les propriétés figurant dans le **Tableau C.14**.

⁽¹⁾ « Association Expérimentique », I.U.T. d'Orléans département OGP, BP 6729, 45067 Orléans Cedex 2

Tableau C.14 Coordonnées des sommets des simplexes initiaux pour 4 facteurs

Ligne Simpl.	X ₁	X ₂	X ₃	X ₄
Ligne Simpl. 1	0,0000	0,0000	0,0000	0,0000
Ligne Simpl. 2	1,0000	0,0000	0,0000	0,0000
Ligne Simpl. 3	0,5000	0,8660	0,0000	0,0000
Ligne Simpl. 4	0,5000	0,2887	0,8165	0,0000
Ligne Simpl. 5	0,5000	0,2887	0,2041	0,7906
Ligne Simpl. 6	0,5000	0,2887	0,2041	0,1581

Les autres lignes de la matrice d'expériences sont obtenues en appliquant des relations de symétrie de façon à couvrir uniformément la circonférence du cercle de rayon unitaire.

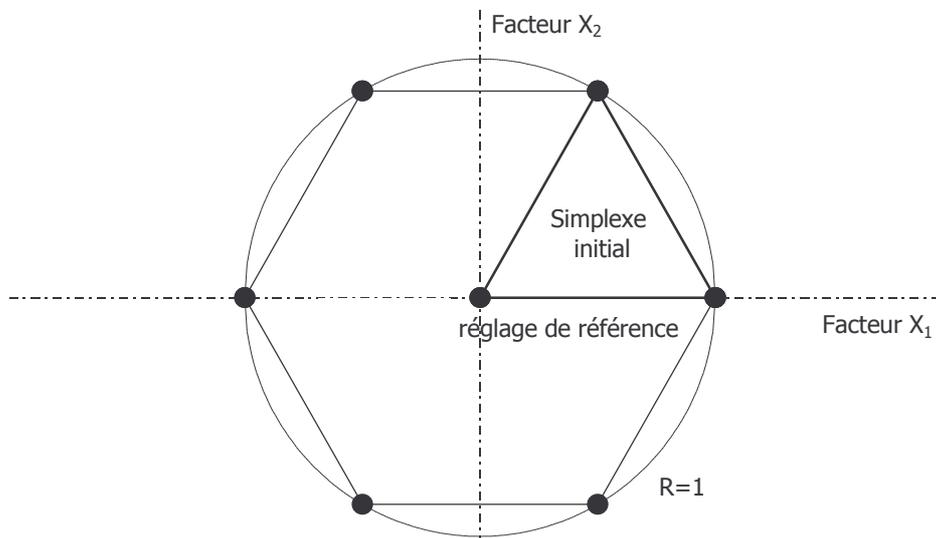


Figure C.11 Réseau uniforme de Doelhart (pour 2 facteurs)

Le **Tableau C.15** représente la matrice d'expériences obtenue pour quatre facteurs. La colonne construction indique comment sont calculées les différentes lignes de la matrice d'expériences :

- les lignes appelées **Lignes Simpl. 1, 2, 3, 4 et 5** indiquent qu'il faut se référer aux lignes correspondantes du **Tableau C.15**,
- les autres constructions indiquent les opérations à faire (voir **Tableau C.16**) pour obtenir les valeurs de la ligne correspondante. Par exemple, pour la ligne 17, il faut faire l'opération (ligne 4) – (ligne 14).

Tableau C.15 Matrice d'expériences associées à un réseau de DOEHLERT pour 4 facteurs

Lignes	X ₁	X ₂	X ₃	X ₄	Construction
1	0,0000	0,0000	0,0000	0,0000	Ligne Simpl. 1
2	1,0000	0,0000	0,0000	0,0000	Ligne Simpl. 2
3	-1,0000	0,0000	0,0000	0,0000	(1)-(2)
4	0,5000	0,8660	0,0000	0,0000	Ligne Simpl. 3
5	-0,5000	-0,8660	0,0000	0,0000	(1)-(4)
6	0,5000	-0,8660	0,0000	0,0000	(2)-(4)
7	-0,5000	0,8660	0,0000	0,0000	(4)-(2)
8	0,5000	0,2887	0,8165	0,0000	Ligne Simpl. 4
9	-0,5000	-0,2887	-0,8165	0,0000	(1)-(8)
10	0,5000	-0,2887	-0,8165	0,0000	(2)-(8)
11	0,0000	0,5774	-0,8165	0,0000	(4)-(8)
12	-0,5000	0,2887	0,8165	0,0000	(8)-(2)
13	0,0000	-0,5774	0,8165	0,0000	(8)-(4)
14	0,5000	0,2887	0,2041	0,7906	Ligne Simpl. 5
15	-0,5000	-0,2887	-0,2041	-0,7906	(1)-(14)
16	0,5000	-0,2887	-0,2041	-0,7906	(2)-(14)
17	0,0000	0,5774	-0,2041	-0,7906	(4)-(14)
18	0,0000	0,0000	0,6124	-0,7906	(8)-(14)
19	-0,5000	0,2887	0,2041	0,7906	(14)-(2)
20	0,0000	-0,5774	0,2041	0,7906	(14)-(4)
21	0,0000	0,0000	-0,6124	0,7906	(14)-(8)

Tableau C.16 Calcul de la ligne 17

	X1	X2	X3	X4
Ligne 4	0,5000	0,8660	0,0000	0,0000
Ligne 14	0,5000	0,2887	0,2041	0,7906

↓

Ligne 17	$0,5 - 0,5 = 0$	$0,8660 - 0,2887$	$0 - 0,2041$	$0 - 0,7906$
(4) - (14)		$= 0,5774$	$= -0,2041$	$= -0,7906$

Avantages des réseaux uniformes de Doehlert :

- La matrice d'expériences est **asymétrique**. Le nombre de niveaux des facteurs est donc différent. C'est un avantage, car les contraintes d'une expérimentation peuvent amener à donner à une ou plusieurs variables le moins de niveaux possible. D'autres facteurs, au contraire, imposent qu'il y ait autant de niveaux que d'expériences.

- Dans notre exemple, on pourra donc affecter :

5 niveaux au facteur X ₁
7 niveaux au facteur X ₂
7 niveaux au facteur X ₃
3 niveaux au facteur X ₄

La multiplication des niveaux de facteur peut représenter un inconvénient pour certains facteurs difficiles à modifier. Dans l'exemple proposé, on affectera ce facteur à X_4 qui ne prend que trois niveaux.

Avec ces matrices, on peut explorer un domaine expérimental voisin, en réutilisant les points du réseau limitrophes de la nouvelle zone à explorer. De plus, ce réseau pourra être resserré, par exemple en divisant la taille de la maille initiale par 2, afin de préciser la connaissance de la zone à explorer.

- Il sera toujours possible d'ajouter de nouveaux facteurs, en faisant peu d'expériences que l'on rajoute au plan (ce qui n'est pas le cas des autres plans d'expériences !). On dit que ces plans sont séquentiels.
- Le nombre d'expériences à faire est peu élevé.

I.3.7. REALISATION DES EXPERIENCES

I.3.7.a. Plan d'expérimentation

Nous allons maintenant utiliser les relations de codage des facteurs pour transformer la matrice d'expériences en plan d'expérimentations. Le **Tableau C.17** représente le plan d'expérimentations à effectuer. Les valeurs expérimentales obtenues lors de la réalisation des films sont légèrement différentes de celles imposées par le plan d'expérimentations (voir **Tableau C.18**). Par conséquent, il faudra rentrer les nouvelles valeurs obtenues dans le plan d'expériences et vérifier que la matrice ainsi reconstruite sera toujours aussi efficace.

N°Exp	Pression	Polarisation	T°	Distance
Unité	Pa	V	°C	cm
1	10	250	450	6
2	1	250	450	6
3	7.8	467	450	6
4	3.3	34	450	6
5	7.8	34	450	6
6	3.3	467	450	6
7	7.8	322	736	6
8	3.3	178	164	6
9	7.8	178	164	6
10	5.5	394	164	6
11	3.3	322	736	6
12	5.5	106	736	6
13	7.8	322	521	7.6
14	3.3	178	379	4.4
15	7.8	178	379	4.4
16	5.5	394	379	4.4
17	5.5	250	664	4.4
18	3.3	322	521	7.6
19	5.5	106	521	7.6
20	5.5	250	236	7.6
21	5.5	250	450	6

Tableau C.17

Plan d'expérimentations

Tableau C.18 Plan d'expérimentations réel

N° expérience	Pression (Pa)	Polarisation (V)	Température (°C)	Distance (cm)
	P	Pola.	T	D
1	9,5	243	460	6
2	1,0	250	455	6
3	7,2	467	460	6
4	3,2	34	500	6
5	7,2	34	465	6
6	3,2	465	465	6
7	7,8	322	770	6
8	3,0	178	164	6
9	8,0	178	164	6
10	5,2	394	164	6
11	3,0	322	736	6
12	5,2	106	736	6
13	8,2	320	526	7,6
14	3,0	178	379	4,4
15	7,8	178	379	4,4
16	5,2	394	395	4,4
17	5,2	250	655	4,4
18	3,1	330	530	7,6
19	5,7	110	535	7,6
20	5,5	250	236	7,6
21	5,2	250	450	6

Nous obtenons la nouvelle matrice d'expériences (**Tableau C.19**) associée aux vraies valeurs du plan d'expérimentations du **Tableau C.18**.

Tableau C.19 Nouvelle matrice d'expériences

N°Exp	X1	X2	X3	X4
1	1	-0,0346	-0,0231	0
2	-1	-0,0023	-0,0396	0
3	0,4588	1	-0,0231	0
4	-0,4824	-1	0,1089	0
5	0,4588	-1	-0,0066	0
6	-0,4824	0,9908	-0,0066	0
7	0,6	0,3303	1	0
8	-0,5294	-0,3349	-1	0
9	0,6471	-0,3349	-1	0
10	-0,0118	0,6628	-1	0
11	-0,5294	0,3303	0,8878	0
12	-0,0118	-0,6674	0,8878	0
13	0,6941	0,321	0,1947	1
14	-0,5294	-0,3349	-0,2904	-1
15	0,6	-0,3349	-0,2904	-1
16	-0,0118	0,6628	-0,2376	-1
17	-0,0118	-0,0023	0,6205	-1
18	-0,5059	0,3672	0,2079	1
19	0,1059	-0,649	0,2244	1
20	0,0588	-0,0023	-0,7624	1
21	-0,0118	-0,0023	-0,0561	0

Pour savoir si la nouvelle matrice d'expériences a conservée sa qualité, on regarde le **facteur d'inflation** $f(b_i)$ (voir **Partie A**). Ce facteur est calculé pour tous les coefficients du modèle, donc chacun de ces coefficients aura une estimation de sa propre représentation (voir **Tableau C.20**).

Tableau C.20 Facteur d'inflation

Coefficient	Fact. Inflation - fi
b1	1.0438
b2	1.0034
b3	1.0480
b4	1.0362
b11	2.3894
b22	2.3895
b33	2.2276
b44	2.4178
b12	1.3219
b13	1.2417
b23	1.2678
b14	1.2028
b24	1.1547
b34	1.1780

On remarque que $1,0034 \leq f_i \leq 2,4117$ pour tous les coefficients, on peut donc dire que la nouvelle matrice apportera l'information désirée car la matrice est de qualité suffisante si $1 \leq f(b_i) \leq 4$.

I.3.7.b. Résultats d'essais

Les valeurs des différentes sorties du plan à surface de réponses sont regroupées dans le **Tableau C.21**.

Tableau C.21 Résultats du plan d'expérimentations

Exp.	dm/St	Vdepot	Vdissol	densité	OH/e	Adh.
Unité	$\mu\text{g.cm}^{-1}\text{.min}^{-1}$	nm.min^{-1}	nm.s^{-1}			N
1	7,3	110,5	218	0,92	59,7	3,5
2	25,9	125,1	1,9	1,73	0	22
3	13,9	70,3	7,8	1,98	18,2	3
4	23,5	121,5	19	1,78	65,3	2
5	15,7	93,8	262,8	1,62	71,6	2
6	23,9	92	1,4	2,66	0	6
7	9,6	59,8	38,3	2,22	35,4	18
8	21,7	112,8	10,9	2,05	24	1
9	20,1	114,5	316,3	1,32	53,9	2
10	19,9	122,5	12	1,09	12,4	1
11	24,6	83,3	0,3	3,26	0,1	8
12	16,6	82,3	22,4	1,67	56,2	3
13	11,9	52,3	36	2,45	33	20
14	30	128,5	2,4	2,22	9,1	1,5
15	22,4	168	108,3	1,93	40,5	5
16	27,5	95,8	2	2,32	22,3	12
17	25,4	130,5	0,8	2,21	6,5	8
18	18,4	82,3	1,4	1,7	0	11
19	12,6	79,5	10,8	1,97	52	3
20	15,7	138,5	131	0,93	41,5	4
21	9,3	99	3,5	0,81	36,9	2

I.3.8. ANALYSE GLOBALE DES RESULTATS D'ESSAIS

Grâce à la **Figure C.12**, on remarque que :

- les résultats expérimentaux des propriétés **dm/St**, **V_{dépot}**, **densité** et **OH/e** présentent une distribution relativement symétrique (médiane \approx moyenne),
- **V_{dissol}** présente une distribution asymétrique (plus de résultats dans les valeurs basses que hautes) avec 3 expériences de forte valeur que l'on peut considérer comme atypiques,
- **Adh** a une distribution de ces résultats également asymétrique, plus de basses valeurs que de fortes valeurs ainsi qu'une valeur atypique ayant une très forte valeur.

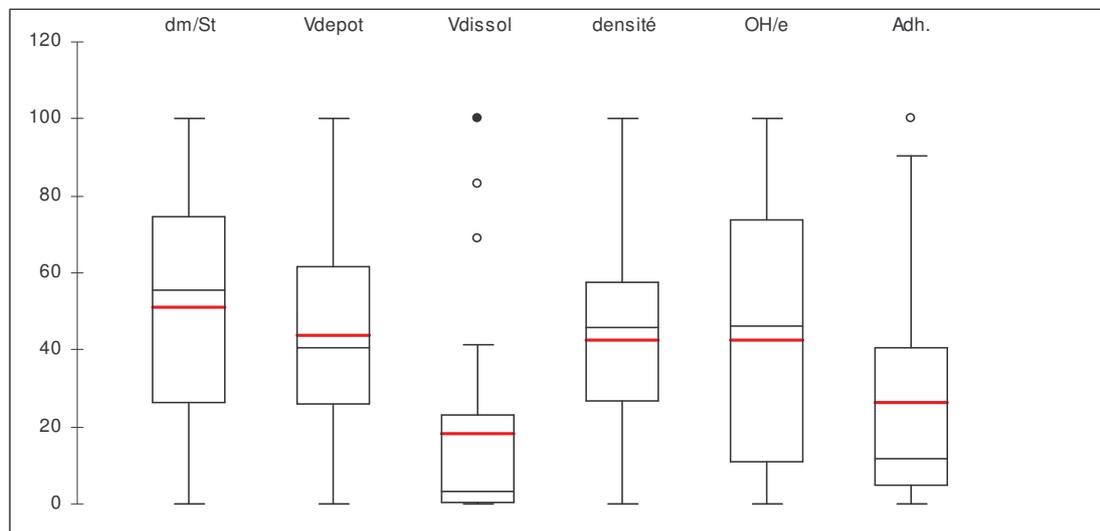


Figure C.12 Box plot

I.3.9. ANALYSE MATHÉMATIQUE DES RESULTATS D'ESSAIS

I.3.9.a. Objectif de l'analyse mathématique

L'analyse mathématique consiste à estimer, grâce à la **méthode des moindres carrés**, les **p** coefficients du modèle et les **N** résidus (écarts entre les valeurs observées et les valeurs prévues par le modèle) pour chacun des traitements du plan d'expériences. La méthode des moindres carrés nécessite un outil informatique adapté ; nous utiliserons dans ce chapitre les résultats obtenus par le logiciel **NemrodW**.

La méthodologie d'analyse des résultats d'essais consiste à établir un modèle mathématique pour chacune des réponses observées

I.3.9.b. Estimation des p coefficients et des N résidus du modèle⁽¹⁾

□ **APPLICATION DE LA METHODE DES MOINDRES CARRÉS :**

Nous ferons ici, un rappel succinct de la méthode des moindres carrés, facile à mettre en œuvre à partir d'un tableur informatique ou d'un logiciel spécialisé.

Pour une réponse donnée et pour le $i^{\text{ème}}$ traitement du plan d'expériences, on peut écrire compte-tenu de la forme générale du modèle (équation 5), la relation suivante :

$$y_i = b_0 + b_1 x_{i,1} + b_2 x_{i,2} + b_3 x_{i,3} + b_4 x_{i,4} + b_{11} x_{i,1}^2 + b_{22} x_{i,2}^2 + b_{33} x_{i,3}^2 + b_{44} x_{i,4}^2 + b_{12} x_{i,1}x_{i,2} + b_{13} x_{i,1}x_{i,3} + b_{23} x_{i,2}x_{i,3} + b_{14} x_{i,1}x_{i,4} + b_{24} x_{i,2}x_{i,4} + b_{34} x_{i,3}x_{i,4} + e_i$$

avec : y_i la valeur observée de la réponse pour le $i^{\text{ème}}$ traitement,

b_i une estimation du coefficient β_i ,

b_{ii} une estimation du coefficient β_{ii} ,

b_{ij} une estimation du coefficient β_{ij} ,

$x_{i,j}$ la valeur que l'on peut lire dans la matrice d'expériences pour le $i^{\text{ème}}$ traitement et pour le $j^{\text{ème}}$ facteur,

et e_i l'écart entre la valeur observée et la valeur calculée.

Nous sommes donc amenés à résoudre le système d'équation ci-dessous :

$$\begin{cases} y_1 = b_0 + b_1 x_{1,1} + \dots + b_{11} x_{1,1}^2 + \dots + b_{12} x_{1,1}x_{1,2} + \dots + b_{34} x_{1,3}x_{1,4} + e_1 \\ \dots \\ y_i = b_0 + b_1 x_{i,1} + \dots + b_{11} x_{i,1}^2 + \dots + b_{12} x_{i,1}x_{i,2} + \dots + b_{34} x_{i,3}x_{i,4} + e_i \\ \dots \\ y_{21} = b_0 + b_1 x_{21,1} + \dots + b_{11} x_{21,1}^2 + \dots + b_{12} x_{21,1}x_{21,2} + \dots + b_{34} x_{21,3}x_{21,4} + e_{21} \end{cases}$$

Si on réécrit ce système sous forme matricielle on obtient :

$$\begin{pmatrix} y_1 \\ \dots \\ y_i \\ \dots \\ y_{21} \end{pmatrix} = \begin{pmatrix} 1 & x_{1,1} & x_{1,2} & \dots & x_{1,1}^2 & \dots & x_{1,1}x_{1,2} & \dots & x_{1,3}x_{1,4} \\ \dots & \dots \\ 1 & x_{i,1} & x_{i,2} & \dots & x_{i,1}^2 & \dots & x_{i,1}x_{i,2} & \dots & x_{i,3}x_{i,4} \\ \dots & \dots \\ 1 & x_{21,1} & x_{21,2} & \dots & x_{21,1}^2 & \dots & x_{21,1}x_{21,2} & \dots & x_{21,3}x_{21,4} \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ \dots \\ b_{11} \\ \dots \\ b_{12} \\ \dots \\ b_{34} \end{pmatrix} + \begin{pmatrix} e_1 \\ \dots \\ e_i \\ \dots \\ e_{21} \end{pmatrix}$$

Si on utilise les notations plus générales, on obtient :

$$(Y) = (X)(B) + (E)$$

avec (X) : matrice du modèle,

(Y) : matrice des résultats d'essais pour une réponse donnée,

(B) : vecteur des estimateurs des coefficients,

(E) : matrice des résidus.

L'estimation des coefficients à partir de la méthode des moindres carrés fait appel à l'équation matricielle suivante : $(B) = (^tXX)^{-1} (^tX)(Y)$.

Les réponses estimées par le modèle sont alors obtenues à partir de l'équation suivante :

$$(\hat{Y}) = (X)(B)$$

Par suite, on obtient donc le résidu pour le traitement i à partir de la relation :

$$e_i = y_i - \hat{y}_i$$

La comparaison des **réponses mesurées** (y_i) et des **réponses calculées** (\hat{y}_i) permet de conclure sur la qualité du modèle.

□ ESTIMATEURS DES COEFFICIENTS

Les estimations des coefficients pour les différents modèles des différentes réponses sont indiquées dans le **Tableau C.22**.

Tableau C.22 Estimation des coefficients des différents modèles liés aux différentes réponses

Coeff.	Vdepot	dm/St	OH/e	densité	Vdissol	Adh
b0	95.744	8.825	35.329	0.860	2.859	0.024
b1	-8.935	-8.781	26.368	-0.411	100.159	-1.832
b2	-15.095	0.143	-28.067	0.276	-59.073	3.777
b3	-23.522	-0.990	-3.090	0.554	-46.167	4.657
b4	-21.138	-5.797	5.482	-0.212	0.264	1.163
b11	21.771	7.967	-5.430	0.511	99.454	13.828
b22	-7.176	8.529	4.875	1.055	51.474	0.106
b33	-3.501	6.942	-6.066	0.796	27.629	2.859
b44	12.638	7.997	-9.622	0.721	-4.796	5.380
b12	2.059	-1.210	7.671	-0.256	-120.425	-0.829
b13	-9.110	-6.231	0.266	-0.168	-73.618	3.905
b23	-19.910	1.087	-4.049	1.020	55.587	0.325
b14	-30.423	2.398	-3.547	0.605	16.129	1.222
b24	13.735	0.442	-13.380	-0.121	2.658	-2.683
b34	-38.977	-0.405	-1.612	0.545	-69.106	0.254

I.3.9.c. Comparaison des réponses observées et des réponses calculées

L'usage des statistiques s'impose lorsque les expérimentateurs ont besoin d'un outil d'aide à la décision. On utilise dans ce cas des tests statistiques dont la finalité consiste à restituer une probabilité, c'est-à-dire un degré de confiance, en regard d'une hypothèse formulée. Quand les premières informations obtenues lors de l'analyse mathématique des résultats d'essais sont suffisamment explicites pour apporter des éléments de réponse aux expérimentateurs, il n'est pas nécessaire de mettre en œuvre, à tout prix, des tests statistiques. C'est dans ce but que nous allons dans un premier temps observer le **Tableau C.23**.

La comparaison des colonnes deux à deux du **Tableau C.23**, nous donne les informations suivantes :

- les modèles des réponses $V_{\text{dépot}}$, **dm/St**, **OH/e** et **densité** bénéficient d'une bonne qualité descriptive, nous pourront donc faire une analyse graphique sur ces réponses,
- les modèles des réponses V_{dissol} et **Adh** ne sont pas à première vue de qualité suffisante, cette différence entre la réponse expérimentale et la réponse calculée nécessitera un test statistique pour savoir si on pourra valider ces modèles.

Tableau C.23 Réponses expérimentales/ Réponses calculées

N°essai	Vdepot		dm/St		OH/e		densité		Adh		Vdissol	
	Yexp	Ycalculé	Yexp	Ycalculé	Yexp	Ycalculé	Yexp	Ycalculé	Yexp	Ycalculé	Yexp	Ycalculé
1	110,5	109,759	7,3	8,229	59,7	57,039	0,92	0,954	3,5	11,723	218	211,577
2	125,1	127,053	25,9	25,373	0	3,737	1,73	1,754	22	15,648	1,9	0,973
3	70,3	76	13,9	14,658	18,2	26,771	1,98	1,959	3	5,442	7,8	7,475
4	121,5	114,075	23,5	22,9	65,3	58,008	1,78	1,8	2	0,355	19	23,259
5	93,8	103,253	15,7	15,447	71,6	75,7	1,62	1,679	2	-1,235	262,8	236,447
6	92	82,392	23,9	23,987	0	-5,293	2,66	2,598	6	8,347	1,4	26,945
7	59,8	53,796	9,6	9,734	35,4	31,644	2,22	2,54	18	14,965	38,3	16,649
8	112,8	119,722	21,7	21,399	24	26,963	2,05	1,696	1	3,847	10,9	35,349
9	114,5	122,13	20,1	19,978	53,9	53,898	1,32	1,582	2	-0,661	316,3	301,001
10	122,5	115,789	19,9	19,92	12,4	18,209	1,09	1,078	1	0,636	12	22,178
11	83,3	75,25	24,6	24,736	0,1	0,932	3,26	2,969	8	10,921	0,3	16,557
12	82,3	90,997	16,6	16,637	56,2	50,855	1,67	1,666	3	3,723	22,4	11,727
13	52,3	54,919	11,9	10,447	33	31,958	2,45	2,137	20	14,597	36	59,448
14	128,5	125,34	30	31,515	9,1	9,255	2,22	2,552	1,5	7,029	2,4	-23,963
15	168	153,554	22,4	22,026	40,5	39,62	1,93	1,597	5	3,715	108,3	148,564
16	95,8	106,233	27,5	26,653	22,3	12,867	2,32	2,415	12	7,688	2	-33,391
17	130,5	137,673	25,4	25,106	6,5	16,658	2,21	2,116	8	8,067	0,8	22,29
18	82,3	97,168	18,4	18,632	0	0,69	1,7	2,006	11	11,743	1,4	-37,24
19	79,5	69,52	12,6	13,464	52	61,204	1,97	1,89	3	7,341	10,8	43,995
20	138,5	130,993	15,7	16,057	41,5	32,648	0,93	1,016	4	4,319	131	112,997
21	99	97,187	9,3	9,002	36,9	35,237	0,81	0,836	2	-0,211	3,5	4,463

Pour porter un jugement sur la qualité descriptive du modèle empirique, on peut tracer un **graphe d'adéquation** du modèle (**Figure C.13**).

Ce graphe restitue le nuage de points défini à partir des valeurs des réponses mesurées et des réponses calculées par le modèle. Plus le nuage de points est aligné sur la première bissectrice, plus la qualité descriptive du modèle est importante.

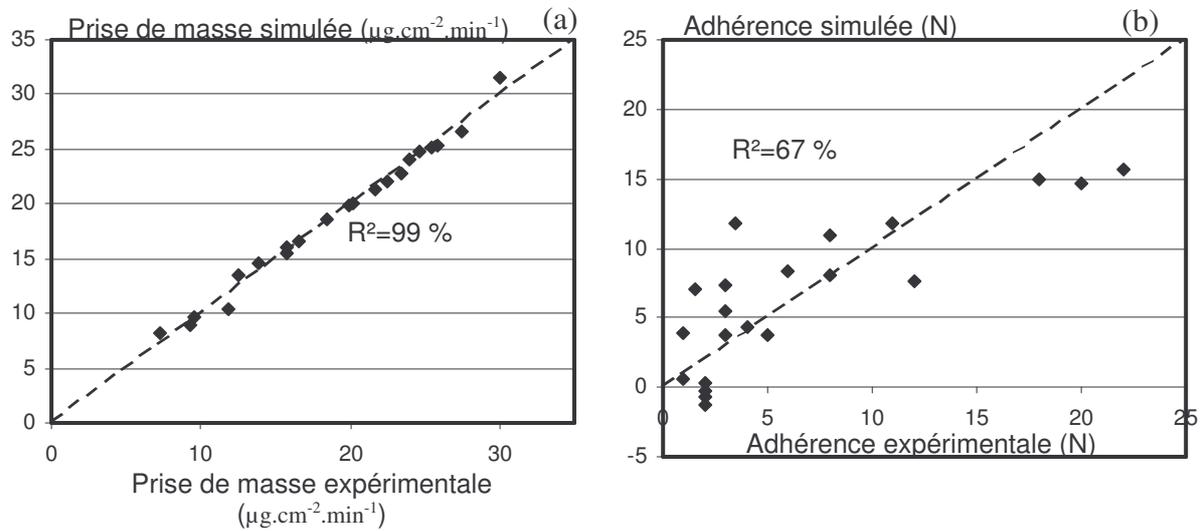


Figure C.13 Graphe d'adéquation pour la prise de masse (a) et l'adhérence (b)

D'après ce graphe, on peut dire que la qualité descriptive du modèle de la prise de masse (dm/St) est importante et que celle de l'adhérence est très mauvaise. La construction d'un tableau d'analyse de régression confirmera ces résultats.

I.4. ANALYSE STATISTIQUE DES RESULTATS

L'équation du modèle empirique n'est qu'une approximation de la réalité (**Partie A**). La mise en œuvre de tests statistiques doit nous permettre de porter un jugement sur les résultats obtenus :

- connaître la qualité des estimations des coefficients,
- estimer la validité du modèle avec les résidus traduisant les écarts entre les valeurs mesurées et les valeurs calculées (**tableau d'analyse de la régression**)
- exprimer la qualité descriptive du modèle dans le domaine expérimental (R_a^2)

A chacun des questionnements correspond un test statistique spécifique (voir **Partie A**) que nous allons appliquer.

I.4.1.a. Qualité des estimateurs des coefficients

On rappelle que l'incertitude sur les coefficients est donnée par la formule (voir **Partie A**) :

$$\delta_i = \pm t \sqrt{\text{var}(b_i)}$$

avec t , la valeur de **Student** au risque 5% qui nous donnera la valeur d'un t_{exp} associé à chacun des coefficients. Cette analyse statistique des coefficients est basée sur l'hypothèse nulle suivante :

H_0 : le coefficient b_i est nul

La probabilité associée à cette hypothèse nous permet d'obtenir le pourcentage associé à l'hypothèse nulle (H_0) (voir **Tableau C.24**).

Tableau C.24 Tableau d'analyse des coefficients

Coeff.	Vdepot	t.exp.	Signif. %
b0	95.744	6,45	0,0865 ***
b1	-8.935	-1,37	21,8
b2	-15.095	-2,57	4,15 *
b3	-23.522	-4	0,741 **
b4	-21.138	-3,89	0,832 **
b11	21.771	1,2	27,4
b22	-7.176	-0,51	63,2
b33	-3.501	-0,29	77,7
b44	12.638	1,2	27,6
b12	2.059	0,13	89,7
b13	-9.110	-0,63	55,4
b23	-19.910	-1,33	23
b14	-30.423	-2,17	7,2
b24	13.735	0,97	37
b34	-38.977	-2,74	3,34 *

On regarde ce tableau pour savoir si on peut choisir un modèle de degré inférieur. Ici, on aurait pu choisir un modèle de degré 1 puisque les coefficients associés à un modèle de degré 2 ont un pourcentage d'être nul élevé (27,4% pour b_{11} , 63,2% pour b_{22} , 77,7% pour b_{33} , 27,6% pour b_{44}).

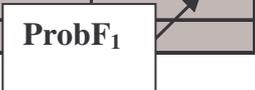
Pour pouvoir rechercher un optimum entre différents critères, ce qui est la finalité de cet exemple et de bon nombre d'études industrielles, certains logiciels de plans d'expériences imposent la description de l'ensemble des réponses avec un même type de modèle. C'est pour cette raison que l'on conservera ici un modèle polynomial de degré 2 pour toutes les réponses.

I.4.1.b. Validité du modèle

L'analyse statistique aboutissant à la validité du modèle est décrite dans un tableau d'analyse de régression (voir Partie A pour les explications du calcul de la probabilité). Nous n'avons pas pu faire de répétitions, donc nous n'avons pas pu faire le test de validation 1 (**Partie A**). Les résultats des test de validation 2 pour les réponses : Adh, Vdissol et dm/St sont respectivement reportés dans les **Tableau C.25**, **Tableau C.26** et **Tableau C.27**.

Tableau C.25 Tableau d'analyse de régression pour la réponse **Adh**

Source de variation	Somme des carrés	ddl	Carré moyen	Rapport	Probabilité
Régression	560,0396	14	40,0028	0,887	60,4
Résidus	270,6032	6	45,1005		
Total	830,6429	20			

ProbF₁


Compte tenu des résultats statistiques obtenus pour l'adhérence (Probabilité du coefficient de Fischer $F_1 = 60\%$: plus de la moitié de la réalité n'est pas représentée par le plan), les résultats graphiques seront à prendre avec beaucoup de précaution concernant cette réponse.

Dans le logiciel **NemrodW**, les étoiles indiquent la qualité du pourcentage obtenu avec le test de **Fischer-Snedecor** (F_1) :

☆☆☆	Prob $F_1 < 0,1\%$
☆☆	Prob $F_1 < 1\%$
☆	Prob $F_1 < 5\%$

Tableau C.26 Tableau d'analyse de régression pour la réponse V_{dissol}

Source de variation	Somme des carrés	ddl	Carré moyen	Rapport	Probabilité
Régression	1,70E+05	14	1,21E+04	6,7962	1,47 *
Résidus	1,07E+04	6	1,78E+03		
Total	1,80E+05	20			

La réponse V_{dissol} a une très faible probabilité F_1 , on valide donc le modèle pour la réponse V_{dissol} . Par ailleurs, le **Tableau C.27** confirme le graphe d'adéquation de la **Figure C.13**.

Tableau C.27 Tableau d'analyse de régression pour la réponse **dm/St**

Source de variation	Somme des carrés	ddl	Carré moyen	Rapport	Probabilité
Régression	833,8146	14	59,5582	41,2751	0,0268 ***
Résidus	8,6577	6	1,443		
Total	842,4724	20			

I.4.1.c. Qualité descriptive du modèle

Le coefficient de détermination traduit la contribution du modèle dans la restitution de la variation de la réponse observée. Par définition, le coefficient de détermination appartient à l'intervalle suivant : $0 \leq R^2 \leq 1$. En présence de plusieurs variables explicatives, il faut éviter l'utilisation du coefficient de détermination R^2 pour comparer la qualité descriptive de différents modèles. Il faut recourir à l'utilisation du coefficient de détermination ajusté $R^2_{\text{ajusté}}$ (**Partie A**).

Rappel : Plus la valeur du coefficient de détermination ajusté est proche de 1 (100%), plus la qualité descriptive du modèle est satisfaisante

Tableau C.28 Qualité descriptive des modèles

Coeff.	Vdepot	dm/St	OH/e	densité	Vdissol	Adh
r^2	91,40%	99,00%	94,20%	88,70%	94,10%	67,40%
r^2_{aj}	80,00%	97,60%	86,50%	73,60%	86,20%	24,00%

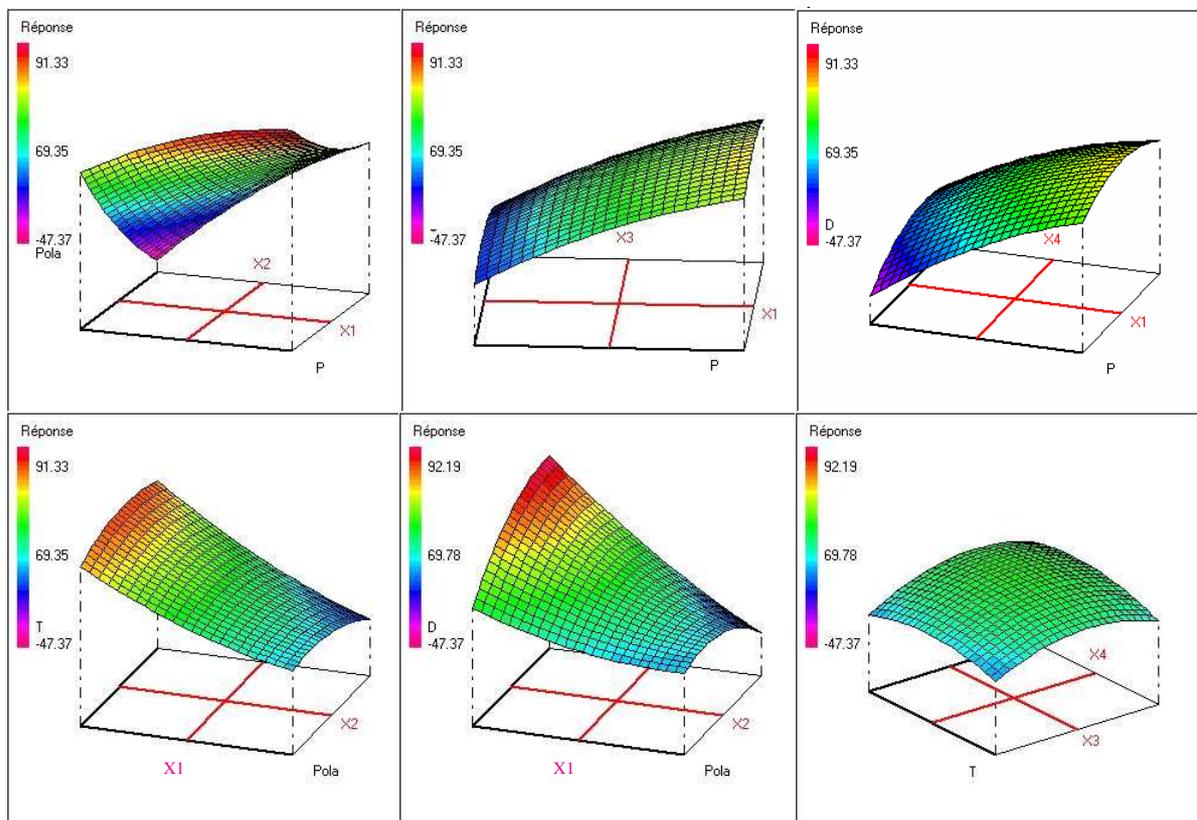
D'après le **Tableau C.28**, on peut dire que toutes les réponses ont une qualité descriptive satisfaisante mise à part la réponse **Adh**.

I.4.2. ANALYSE GRAPHIQUE DES RESULTATS

Le domaine expérimental étant défini à partir de la variation de quatre facteurs, il est difficile de restituer de façon simple la variation de la réponse. On a donc recours à des coupes et des projections qui consistent à fixer un niveau donné à certains facteurs. Les espaces de représentation des surfaces de réponse sont définis à partir de la variation de deux facteurs. Nous utiliserons le logiciel **NemrodW** pour la construction de ces graphiques. L'analyse graphique du modèle consiste à restituer l'équation de ce dernier sous deux formes : les surfaces de réponse et les courbes iso-réponses

I.4.2.a. Surfaces de réponses

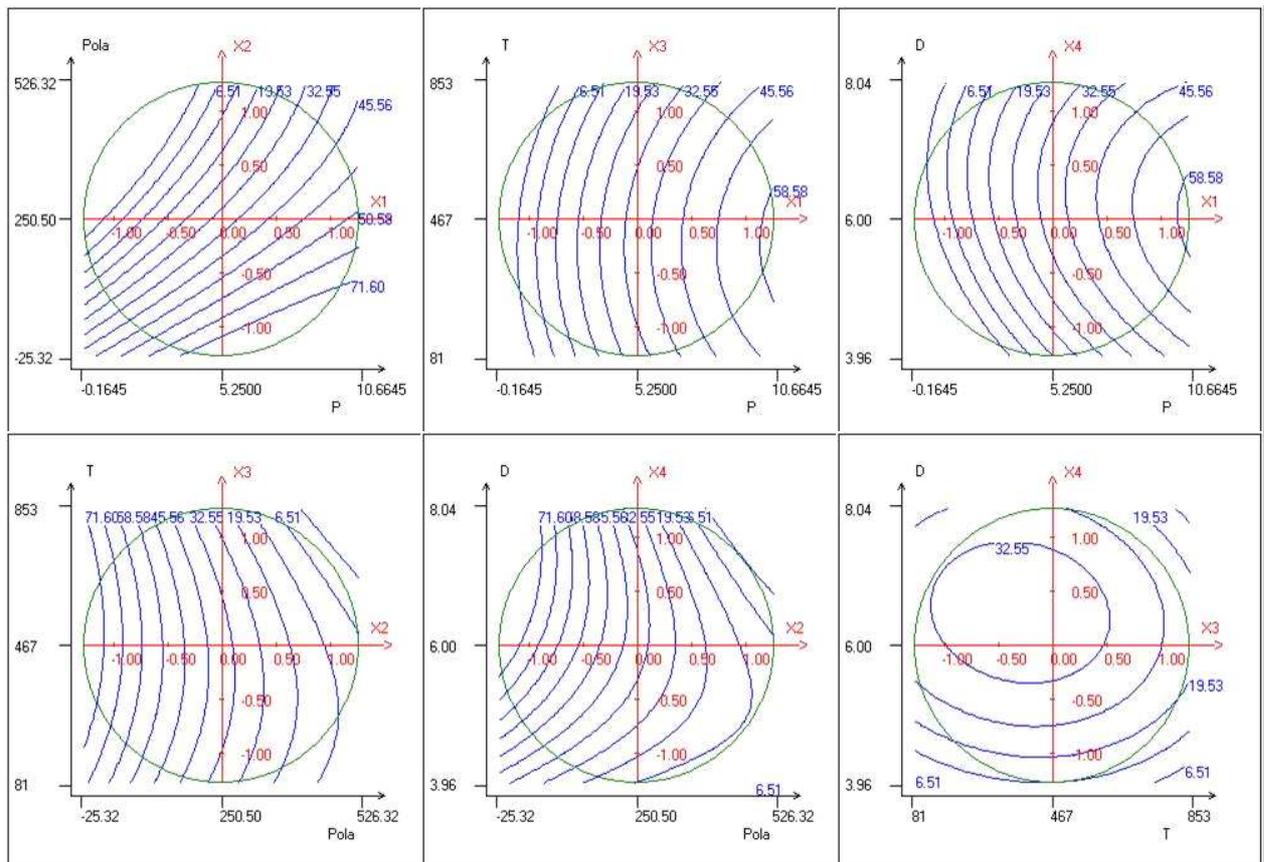
Les surfaces de réponse sont des graphiques en trois dimensions (voir **Figure C.14**). Le plan horizontal de la figure matérialise le domaine de variation de deux facteurs ; l'axe vertical matérialise la variation de la réponse à partir du modèle. Au-delà de deux facteurs, il est nécessaire de maintenir à un niveau constant les facteurs dont les variations ne sont pas décrites dans le plan horizontal. Nous allons nous intéresser à l'étude de la réponse **OH/e**.



X1 : P (Pression)
X2 : Pola (Polarisation)
X3 : T (Température)
X4 : d (distance)

Figure C.14 Surface de réponse de OH/e**I.4.2.b. Courbes iso-réponses**

Les courbes d'iso-réponses constituent une projection de la surface de réponse dans le plan horizontal (**Figure C.15**). Elles s'interprètent comme des courbes de niveaux sur lesquelles est projetée la valeur de la réponse. Au-delà de deux facteurs, il est nécessaire de maintenir à un niveau constant les facteurs dont les variations ne sont pas décrites dans le plan horizontal.

**Figure C.15** Courbes d'iso-réponses de OH/e**I.4.2.c. Interprétation des surfaces de réponses et des courbes d'iso-réponses pour la réponse OH/e**

- Les facteurs principaux influant sur la quantité d'OH dans le film sont la pression (X1) et la polarisation (X2). Une diminution de la pression dans le réacteur entraîne une forte diminution de la quantité d'OH incorporée dans le film tandis qu'une augmentation de la polarisation conduit au même résultat. Ces évolutions sont plus marquées à faible température et longue distance.

- Les évolutions de la quantité d'OH présent dans le film induites par la modification de la distance et de la température sont plus faibles que pour la pression et la polarisation.

Mais à faible polarisation (inférieure à 250V), la quantité d'OH augmente lorsque la distance injecteur/substrat croît tandis que la température ne permet une diminution des OH qu'à longue distance.

Nous ne montrerons pas ici tous les graphiques correspondant à toutes les réponses, nous allons seulement exposer les conclusions que nous avons tirées de ces graphiques.

I.4.2.d. Conclusion sur les analyses graphiques

Compte tenu du nombre important des réponses analysées, nous n'avons pas jugé utile de reporter l'ensemble des graphiques. A titre indicatif, l'analyse de l'ensemble des résultats obtenus à partir des surfaces de réponses et des courbes d'iso-réponses est résumée ci-dessous [1].

- Le maximum d'**OH** dans le film sera obtenu dans les conditions de forte distance (8 cm), sans polarisation (0V), à forte pression et basse température.
- Le maximum de **vitesse de dissolution** du film est obtenu pour des distances relativement élevées (6 à 8 cm), sans polarisation, à forte pression et basse température.
- une faible **densité** de couche sera obtenue à forte pression et basse température, en polarisant la couche pour des distances injecteur/substrat supérieures à 4 cm.
- Le maximum de **prise de masse** est obtenu lorsque les effets d'une pression faible (1 Pa) et d'une courte distance (4 cm) sont cumulés. Dans ces conditions, une forte température (800°C) augmente légèrement la prise de masse alors que la polarisation a peu d'influence.
- Le minimum de **vitesse de dépôt** est facilement identifié à distance et température élevée (respectivement 8 cm et 800°C), pour des pressions supérieures à 1 Pa sans que la polarisation ait d'influence majeure. Les maxima, quant à eux, dépendent essentiellement de la distance injecteur/substrat. Le premier est observé à 4 cm, sans polarisation, à haute température (800°C) quelle que soit la pression (1-10 Pa). Si on fixe la pression à 10 Pa, toujours à 4 cm, une forte vitesse est observée dans un domaine plus large de polarisation (0-250 V) et de température (450-800 °C). Le second maximum est observé dans des conditions inverses en terme de distance (8 cm) et de température (100 °C), pour une pression obligatoirement basse (1 Pa) avec une polarisation (100-500 V) même si l'influence de celle-ci est limitée. Pour des distances intermédiaires, les variations de vitesses de dépôt sont beaucoup moins marquées.

I.4.3. CONCLUSION DU PLAN D'OPTIMISATION

I.4.3.a. Conclusion des experts en dépôts

L'analyse des résultats du plan nous montre que le procédé de dépôt permet la réalisation de couches minces d'alumine ayant de bonnes caractéristiques dans des conditions très différentes :

- si la pression est fixée à 1 Pa, il existe une condition commune aux trois propriétés qui consiste à ne pas polariser la couche (0 V), à la chauffer au minimum (100 °C) et à se rapprocher de l'injecteur de TMA (4 cm),
- l'autre possibilité est indépendante de la pression (1 à 10 Pa) mais, en revanche, elle nécessite des conditions sévères en terme de polarisation (500 V) et de température (800 °C) pour une distance élevée (8 cm). Ce compromis correspond à une vitesse de dissolution très faible mais non minimale.

De plus, la modification des quatre paramètres d'entrée étudiés (pression, distance, température, polarisation) entraîne des variations importantes des cinétiques de dépôt (prise de masse et vitesse de dépôt).

Le plan de criblage nous a permis de classer les facteurs par ordre de leur importance sur les propriétés physicochimiques et les vitesses de dépôt des films. Le facteur le plus influent est le débit de TMA mais n'a qu'une influence limitée sur les propriétés physico-chimiques des films tandis que le débit d'oxygène et la puissance micro-onde n'ont aucun effet sur les dépôts. Les paramètres les plus intéressants pour une étude ultérieure sont, d'après le plan de criblage et l'expérience des expérimentateurs : la pression dans l'enceinte, la distance injecteur/substrat, la température du substrat et la polarisation radiofréquence. Ces quatre paramètres ont ensuite été utilisés pour construire un plan de surface de réponses prédictif alors que les autres paramètres (utilisés dans le plan de criblage) ont été fixés. Les marqueurs de qualité (vitesse de dissolution, taux d'OH et densité du film) ont des variations très différentes suivant les paramètres expérimentaux. Les conditions optimales pour obtenir des couches de bonnes qualités sont de deux types : un cas à basse pression et un autre dans des conditions plus sévères (500V/800°C) pour le substrat quelle que soit la pression. Les dépôts de mauvaise qualité, au contraire, sont toujours réalisés dans les conditions qui fournissent le moins d'énergie au système (basse température, basse polarisation, haute pression et longue distance).

I.4.3.b. Conclusion sur la stratégie des plans d'expériences retenue⁽¹⁾

Quels sont les avantages ?

- Les expérimentateurs disposent d'un fil conducteur, reproductible d'une étude à l'autre, qui ne demande qu'à être adapté en fonction de la spécificité des problèmes,
- l'approche séquentielle offerte par les réseaux uniformes de **Doelhart**, au même titre que les possibilités de glissement du domaine expérimental, peuvent représenter un avantage,
- la construction de la matrice d'expériences repose sur des considérations géométriques simples ce qui est rare dans le domaine des plans d'expériences pour l'étude des surfaces de réponse,
- le même plan d'expériences est utilisé pour l'étude de l'ensemble des réponses.

Quels sont les inconvénients ?

- Le principal inconvénient des réseaux uniformes de **Doelhart** est qu'ils sont peu répandus dans les logiciels de construction de plans d'expériences (**NemrodW** l'utilise),
- le nombre important de niveaux qu'il est nécessaire d'affecter à certains facteurs peut s'avérer être un inconvénient,
- le nombre d'expériences à réaliser pour les plans à surfaces de réponses croît rapidement en fonction du nombre de facteurs. Il est donc nécessaire de limiter le nombre de facteurs d'une étude d'optimisation à quatre ou cinq tout au plus,
- l'erreur expérimentale n'a pas été calculée et donc, le test de validation du modèle F_1 n'a pu être obtenu par manque de répétitions des expériences au centre du domaine.

I.5. ANALYSES DE DONNEES

Nous allons maintenant analyser les données obtenues à l'aide des deux plans d'expériences (criblage et optimisation) vu précédemment. Nous allons découper cette analyse en deux analyses que nous appellerons : Analyse 1 et Analyse 2.

- **L'analyse 1** va nous permettre de corrélérer les facteurs de sorties entre eux (prise de masse, vitesse d'attaque, densité, adhérence, OH/e) afin d'aider l'expérimentateur à comprendre la structure des films réalisés. Les propriétés des films seront corrélées à la morphologie des splats obtenus par projection plasma.

- **L'analyse 2** va nous permettre de corrélérer les propriétés des films aux données du plasma (spectrométrie d'émission optique) afin d'identifier le chemin réactionnel menant à la construction des dépôts.

Par souci de présentation et d'équilibre entre les différents chapitres de cette partie, nous n'allons pas décrire toutes les étapes (pré-traitement des données, analyses bidimensionnelles, construction des axes ...) de l'analyse multidimensionnelle de données pour toutes les analyses. Mais, n'oublions pas que ces étapes sont nécessaires pour pouvoir faire de l'analyse multidimensionnelle de données.

I.5.1. ANALYSE 1 : PROPRIETES DES FILMS

Avant de faire l'A.C.P. nous allons d'abord établir le tableau de données et voir s'il n'existe pas d'individus atypiques à sortir de l'analyse multidimensionnelle de données.

I.5.1.a. Tableau de données

Le **Tableau C.29** regroupe l'ensemble des réponses des plans d'expériences précédents ainsi que des données déjà existantes sur ces réponses.

Tableau C.29 Tableau de données brutes de l'Analyse 1

N°individu	dm/St	Vdépôt	densité	OH/e	Vdissol	Adh
1	7,3	110,5	0,9	59,7	218,0	3,5
2	25,9	125,1	1,7	0,0	1,9	22,0
3	13,9	70,3	2,0	18,2	7,8	3,0
4	23,5	121,5	1,8	65,3	19,0	2,0
5	15,7	93,8	1,6	71,6	262,8	2,0
6	23,9	92,0	2,7	0,0	1,4	6,0
7	9,6	59,8	2,2	35,4	38,3	18,0
8	21,7	112,8	2,1	24,0	10,9	1,0
9	20,1	114,5	1,3	53,9	316,3	2,0
10	19,9	122,5	1,1	12,4	12,0	1,0
11	24,6	83,3	3,3	0,1	0,3	8,0
12	16,6	82,3	1,7	56,2	22,4	3,0
13	11,9	52,3	2,4	33,0	36,0	20,0
14	30,0	128,5	2,2	9,1	2,4	1,5
15	22,4	168,0	1,9	40,5	108,3	5,0
16	27,5	95,8	2,3	22,3	2,0	12,0
17	25,4	130,5	2,2	6,5	0,8	8,0
18	18,4	82,3	1,7	0,0	1,4	11,0
19	12,6	79,5	2,0	52,0	10,8	3,0
20	15,7	138,5	0,9	41,5	131,0	4,0
21	9,3	99,0	0,8	36,9	3,5	2,0
24	51,5	237,0	2,3	57,0	16,2	15,0
25	47,9	173,0	2,5	75,9	89,0	18,0
26	59,8	216,0	2,5	68,4	9,5	14,0
27	49,7	160,0	3,2	127,5	6,8	19,0
28	46,7	199,0	2,0	43,4	17,0	1,0
29	39,2	164,0	2,5	83,4	329,4	0,5
30	46,1	132,0	2,9	40,3	1,3	18,0
31	33,1	140,0	2,6	2,0	0,4	18,0
32	40,4	158,0	2,4	83,8	304,0	3,0
34	49,3	142,0	3,5	11,3	1,7	16,0
35	34,4	139,0	2,5	97,8	22,0	2,0
36	29,2	103,0	2,7	63,1	6,8	15,0
37	19,5	90,0	2,2	111,8	423,0	2,0
39	27,7	113,0	2,7	111,0	3,2	17,0
38	31,5	225,0	1,5	82,0	1100,0	0,2
23	79,2	300,0	2,8	100,5	13,4	20,0
22	94,7	370,0	2,6	95,5	22,4	9,0
33	34,9	224,0	1,7	79,5	594,0	0,5

A cette étape de récupération de données, il est important de rajouter le plus possible d'expériences existantes de manière à améliorer la qualité de l'analyse multidimensionnelle de données.

I.5.1.b. Pré-traitement des données

Nous rappelons ici que le pré-traitement des données consiste de manière générale à détecter les valeurs suspectes du tableau de données brutes. Pour cela, nous allons tracer les boîtes à moustaches des variables sur un même schéma (voir **Figure C.16**).

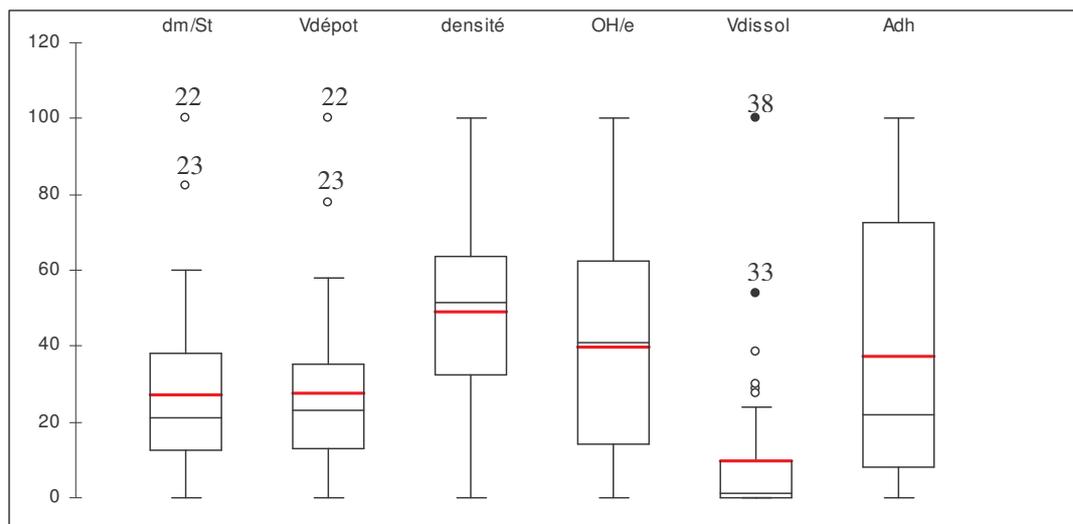


Figure C.16 Box Plot de l'Analyse 1

Les expériences 22 et 23 sont des observations atypiques, il leur correspond des valeurs très élevées de la vitesse de dépôt et de dm/St. Les expériences 38 et 33 sont des expériences atypiques pour la variable V_{dissol} .

Nous retirerons donc de l'analyse, les individus 22, 23, 38 et 33 et nous les déclarerons en tant qu'individus supplémentaires dans l'A.C.P., de manière à positionner ces expériences sur les graphiques obtenus sans qu'elles interviennent dans la construction des axes (voir **Partie A**).

I.5.1.c. Matrice des corrélations et matrice XY

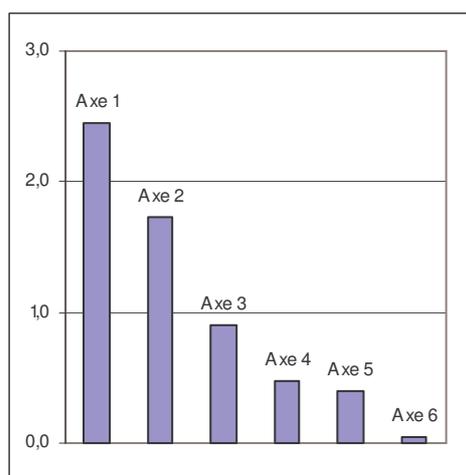
L'analyse bidimensionnelle détermine les relations des variables deux à deux grâce à la matrice des corrélations. Le **Tableau C.30** répertorie toutes les valeurs des coefficients de corrélations entre les six variables. Les coefficients significatifs (positifs ou négatifs) sont exprimés en gras dans le tableau.

Tableau C.30 Matrice des corrélations

	dm/St	Vdépôt	densité	OH/e	Vdissol	Adh
dm/St	1					
Vdépôt	0,817	1				
densité	0,618	0,158	1			
OH/e	0,260	0,265	0,115	1		
Vdissol	-0,125	0,025	-0,234	0,469	1	
Adh	0,358	0,083	130,522	-0,055	-0,405	1

A titre d'exemple, la vitesse de dépôt est fortement corrélée à la prise de masse (dm/St), le coefficient de corrélation étant égal à 0,817. Ces représentations bidimensionnelles des corrélations sont juste des appuis aux résultats de l'A.C.P.. Deux à deux nous devons analyser ici 15 corrélations, 15 graphiques bidimensionnels sont donc nécessaires à l'analyse de ces corrélations. L'intérêt de l'A.C.P. est de faciliter l'analyse graphique des résultats, ainsi le cercle des corrélations remplace en un seul graphique, les 15 graphiques que l'on devrait analyser !

I.5.1.d. Part d'information portée par les axes



L'histogramme des valeurs propres (**Figure C.17**), nous indique qu'il suffit de regarder les résultats sur deux axes (1 et 2). En effet, ce sont les seuls à dépasser la valeur 1 de la valeur propre (voir Partie B). Les valeurs propres λ , le pourcentage de variance et la variance cumulée pour chaque axe sont reportés dans le **Tableau C.31**.

Figure C.17 Histogramme des valeurs propres

Tableau C.31 Part d'information apporté par les axes

	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5	Axe 6
Valeur propre	2,446	1,731	0,905	0,477	0,394	0,048
% variance	40,761	28,850	15,085	7,952	6,559	0,794
% cumulé	40,761	69,611	84,695	92,647	99,206	100,000

On voit ici que les axes 1 et 2 expliquent environ 70% de l'information contenue dans le **Tableau C.29**.

Nous avons donc choisi de regarder les projections des individus et des variables dans le plan défini par les axes 1 et 2. Ainsi, nous passerons d'un espace à 6 dimensions (6 variables à analyser) à un espace à deux dimensions.

Plan principal choisi : Axe 1 / Axe 2

Nous allons regarder maintenant la contribution des individus à la construction des axes 1 et 2.

I.5.1.e. Contribution des individus au calcul des axes

D'après les histogrammes des contributions des individus au calcul des axes 1 et 2 (**Figure C.18**), nous pouvons dire que la construction des axes 1 et 2 est due à un ensemble assez homogène d'individus ; les axes seront qualifiés de « stable » et on leur accordera une certaine signification.

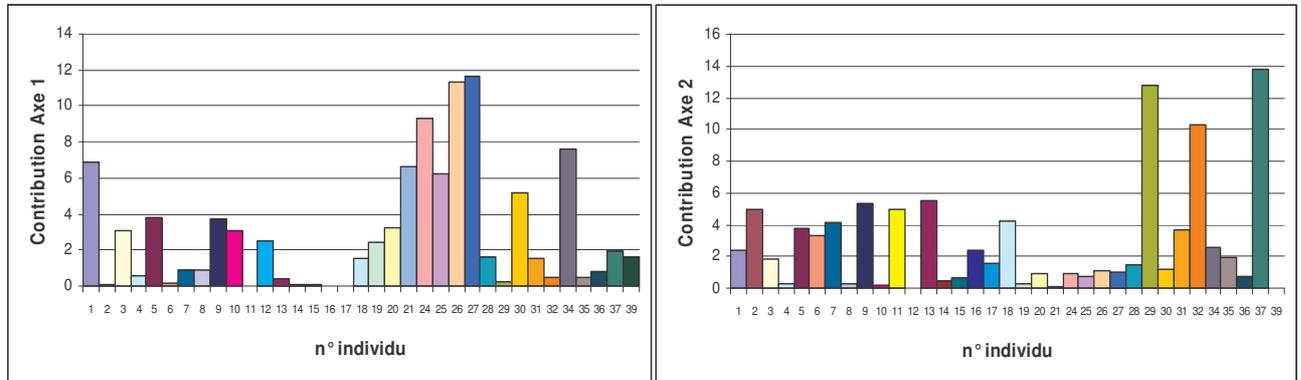


Figure C.18 Contribution des individus au calcul des axes 1 et 2

Avant de regarder les résultats graphiques, nous allons vérifier la qualité de la représentation des individus dans le plan principal.

I.5.1.f. Qualité de représentation des individus

Les plans principaux de l'A.C.P. ne permettent pas d'avoir une vue parfaite de la réalité, il faut donc se méfier des déformations car le pourcentage d'information cumulé n'est pas égal à 100 %. Nous avons vu précédemment que dans cet exemple, le pourcentage d'information cumulé pour les axes 1 et 2 est égal à 70%. Le **Tableau C.32** rappelle les conditions d'une bonne ou d'une mauvaise qualité de représentation des individus dans le plan principal.

Tableau C.32 Qualité de représentation des individus

Valeur de $(\text{Cos } \alpha)^2$	Signification
Autour de 1	Très bonne représentation
Supérieur à 0,5	Représentation acceptable
Inférieur à 0,5	Mauvaise représentation

L'histogramme de la **Figure C.19** représente la somme des cosinus carrés sur les axes 1 et 2 (voir **Partie B**) et donne l'information sur la qualité de représentation des individus. Ainsi, il ne faudra pas interpréter les résultats concernant les expériences 4, 8, 14, 15, 28, 35, 36 et 39 à cause de leur mauvaise représentation dans le plan principal choisi.

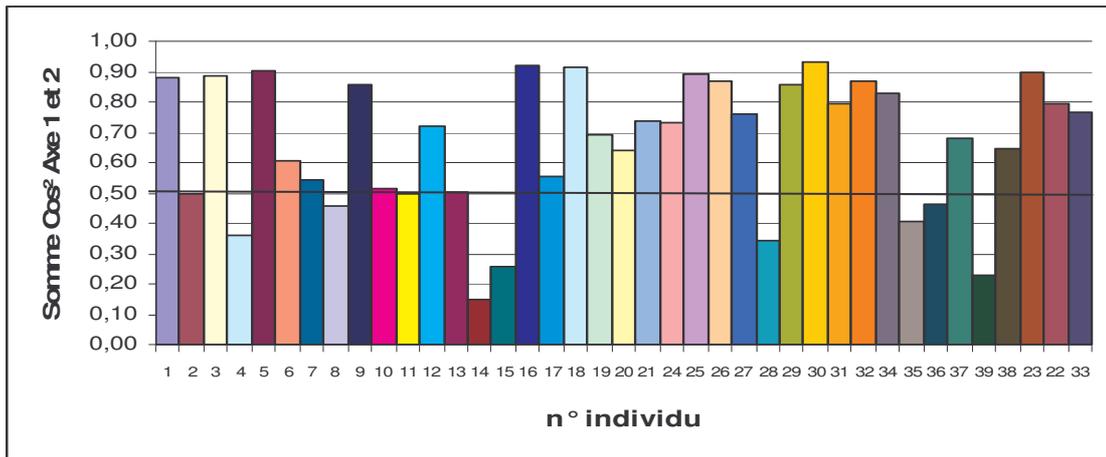


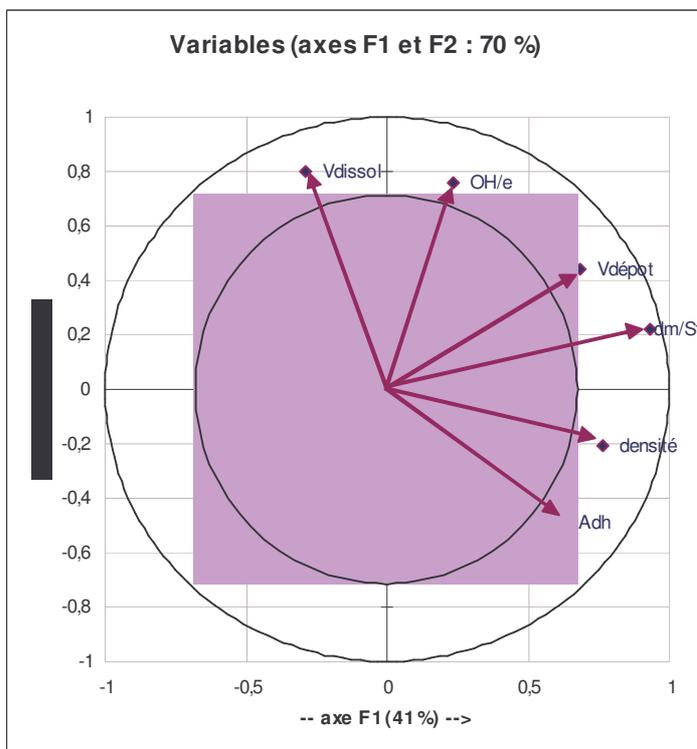
Figure C.19 Qualité de représentation des individus dans le plan principal

Nous pouvons à présent interpréter les résultats de l'analyse graphique.

I.5.1.g. Analyse graphique

 **Cercle des corrélations**

Ce cercle (**Figure C.20**) représente les **corrélations** entre les variables initiales V_{dissol} , OH/e, $V_{dépot}$, dm/St, densité et Adh (Voir **Partie B**).



Analyse mathématique du graphique

- V_{dissol} et OH/e sont positivement corrélées
- V_{dissol} et dm/St sont indépendantes
- V_{dissol} et Adh sont inversement corrélées
- OH/e et $V_{dépot}$ sont positivement corrélées
- OH/e et densité sont indépendantes
- $V_{dépot}$ et dm/St sont positivement corrélées
- dm/St et densité sont positivement corrélées
- densité et Adh sont positivement corrélées

Figure C.20 Cercle des corrélations de l'Analyse 1

Récapitulatif des interprétations physiques :

L'analyse du graphique révèle les corrélations suivantes entre les propriétés des films :

- la prise de masse est liée à la vitesse de dépôt,
- la vitesse de dépôt est liée au taux d'impuretés (OH/e) dans le film,
- le taux d'OH influe sur la vitesse de dissolution dans l'acide fluorhydrique,
- l'augmentation de la densité permet d'accroître la charge critique de rupture du film,
- l'adhérence du film est inversement corrélée avec la dissolution donc de faibles valeurs de cette dernière favorisent l'adhérence du film,
- l'adhérence est liée à la prise de masse par l'intermédiaire de la densité,
- les paramètres qui semblent totalement "décorrélés" sont les suivants : OH/e et vitesse avec l'adhérence, vitesse de dissolution avec prise de masse, densité et OH/e,
- pour les autres paramètres, il n'est pas possible de se prononcer (coefficients de corrélation compris entre 0,1 et 0,32).

Les variations des quatre paramètres expérimentaux (facteurs du plan d'optimisation) dans l'intervalle considéré, modifient donc de façon importante les propriétés des films. Ces différences caractéristiques étant connues, quelques particules d'alumine (splats) ont été projetées sur les films déposés sur un substrat en silicium afin d'observer leur morphologie [2].

Biplot

Dans cette étude, menée en collaboration avec **T. Haure** [2], nous allons nous servir du biplot pour observer les corrélations entre les propriétés des films et la morphologie des splats.

Le graphique de la **Figure C.21** permet de noter les évolutions majeures des morphologies en fonction des paramètres de dépôts PECVD.

Les splats notés 18, 11 et 30, dans la partie basse, présentent une morphologie quasi normale en forme de disque microfissuré. Sur la partie droite, échantillons 22 et 23, la matière se retrouve uniquement sur la bordure. Sur la partie en haut à gauche, les splats ne présentent en effet plus de morphologie circulaire dans cette zone, l'étalement semble influencé par les propriétés des films. On note sur les splats n° 9 et 33, deux zones différentes : la première, au centre, semble marquer l'impact de la particule fondue et la seconde, autour de la première, montre un étalement

inhomogène de la particule. Dans la zone à gauche, sur les splats réalisés dans les conditions n°21, on observe un défaut important d'adhérence des films PECVD entraînant un décollement du film sous le splat. De même, le splat n°18 malgré une morphologie « normale », présente, compte tenu de la zone importante décollée en son centre, une rupture de la couche PECVD. En observant le sens et la direction des flèches montrant l'évolution de chaque caractéristique, il est possible de déterminer les paramètres ayant une influence importante sur la morphologie des splats, en sachant que ces propriétés sont aussi liées entre elles. Il s'avère donc qu'une densité élevée et une faible vitesse de dissolution semblent être les paramètres principaux. Les autres propriétés influant aussi, mais de façon moindre, sur la morphologie des splats. Une densité élevée permet de limiter la formation de pores dans les splats et une vitesse de dissolution importante semble jouer directement sur l'étalement des lamelles [2].

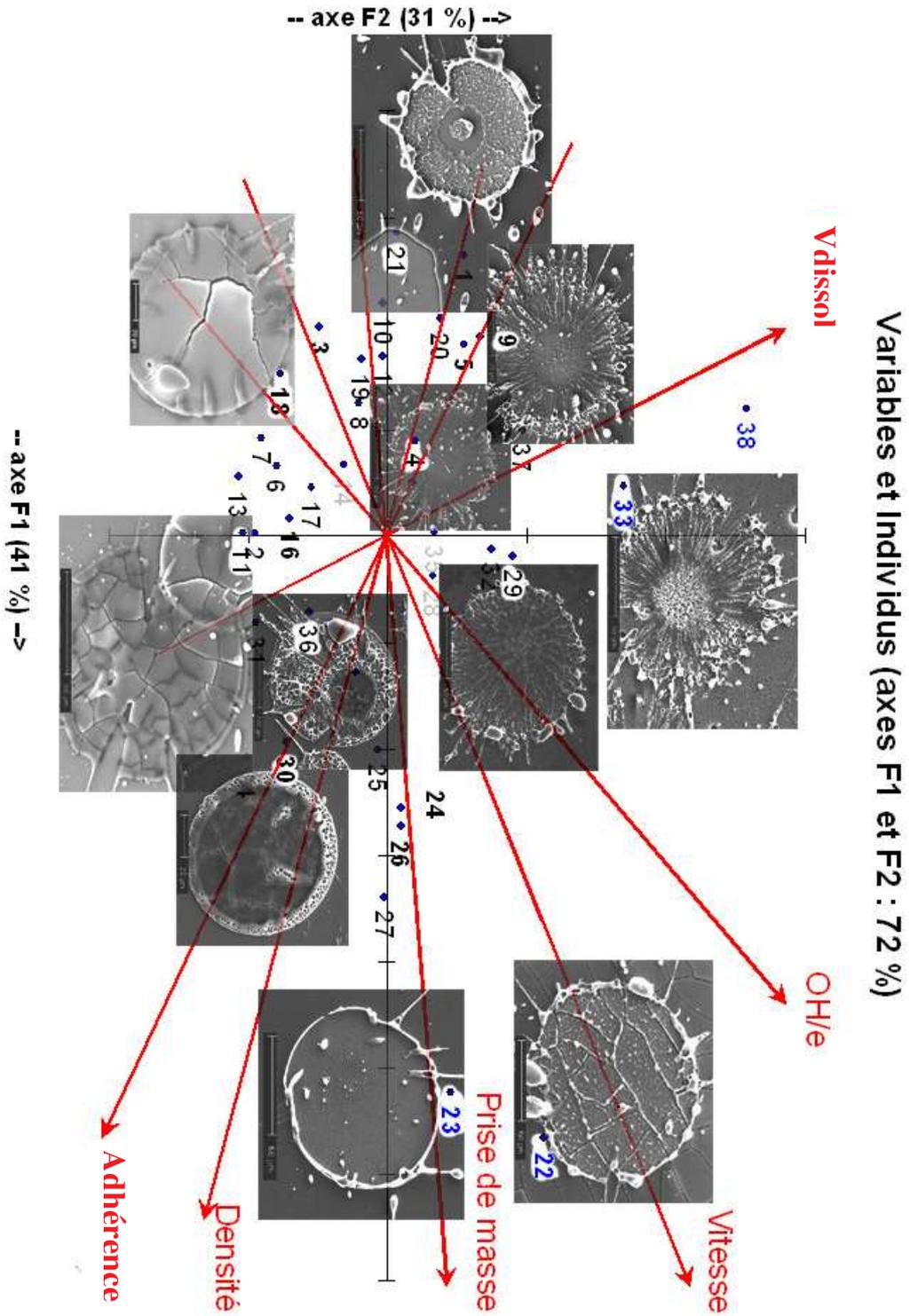


Figure C.21 Biplot : morphologie des plats en fonction des propriétés des sous-couches d'alumine PECVD

I.5.2. ANALYSE 2 : PROPRIETES DES FILMS ET PLASMA

Cette analyse va nous permettre de corrélérer les propriétés des films aux données du plasma (Spectrométrie d'Emission Optique) afin d'identifier le chemin réactionnel menant à la construction des dépôts. Nous ne détaillerons pas ici les étapes de l'analyse. Nous passons donc directement à l'analyse graphique, plus précisément à celle qui nous intéresse : les cercles des corrélations.

I.5.2.a. Influence des paramètres de dépôt sur l'intensités des raies

Nous avons étudié ici l'influence des paramètres d'entrée sur l'intensité des raies obtenues par Spectrométrie d'Emission Optique [1] La Figure C.22 représente la projection sur les deux axes principaux des vecteurs des 3 facteurs (température, pression, polarisation) et de 7 intensités de raies : 83% de l'information est représentée sur ce graphique.

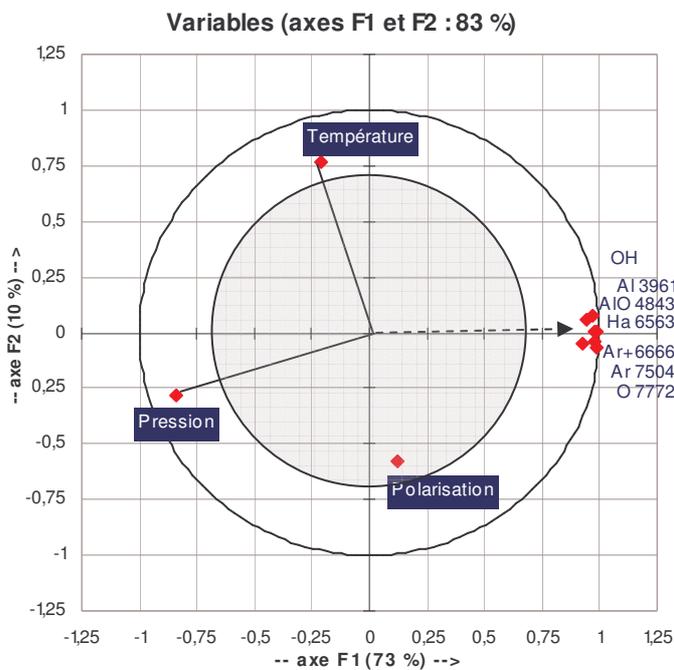


Figure C.22 Cercle des corrélations : relations entre les intensités des raies et les 3 facteurs (projection dans le plan Axe1/Axe2)

L'ensemble des projections des intensités se trouve regroupé à la même position et présente une évolution identique : l'opposition avec le vecteur pression et la perpendicularité avec le vecteur température, indique que seule la pression influe sur l'intensité des raies.

I.5.2.b. Influence des paramètres de dépôt sur les composants du plasma

Les cercles des corrélations de l'A.C.P. (**Figure C.23**) permettent de faire la comparaison entre les 3 paramètres expérimentaux (pression, température, polarisation) et les 6 espèces observables dans la post-décharge (Al, AlO, H, OH, O, O₂⁺).

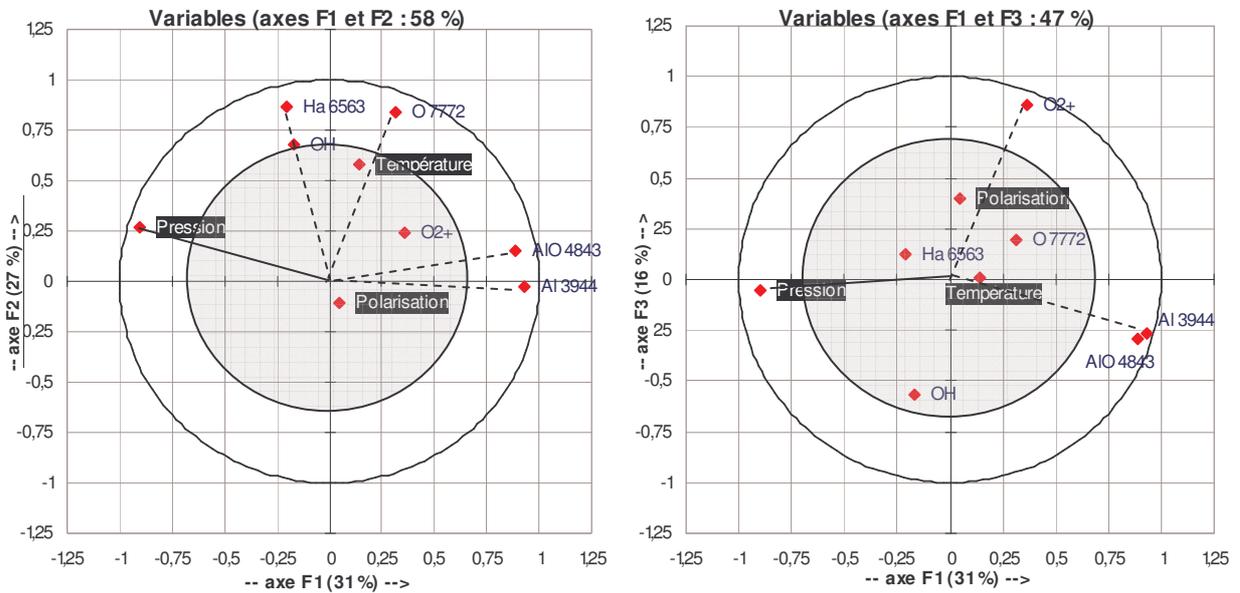


Figure C.23 Relations entre la composition du plasma et les paramètres expérimentaux

Le nombre élevé de variables entraîne une faible représentativité (74% en trois axes) : dans un premier temps l'étude sera limitée à l'influence des 3 paramètres sur les 6 espèces. Dans cet espace à 3 dimensions, seule la pression peut être bien représentée (les deux autres ont des projections inférieures 0,7).

La pression a un effet sensible sur la « quantité » d'espèces contenant de l'aluminium dans le plasma (Al, AlO) : plus la pression est élevée, moins ces espèces peuvent être détectées.

I.5.2.c. Corrélation entre les espèces du plasma

Compte tenu de la faible représentativité des cercles précédemment cités, nous avons retiré les paramètres expérimentaux afin d'obtenir une meilleure représentation des projections. Le retrait de ces 3 variables permet de représenter 92% de l'information grâce aux 3 premiers axes principaux.

La **Figure C.24** représente le tableau des corrélations et les projections sur les trois axes principaux (les plus représentatifs), des « concentrations » en espèces détectées dans le réacteur à 2 cm au-dessous de l'injecteur.

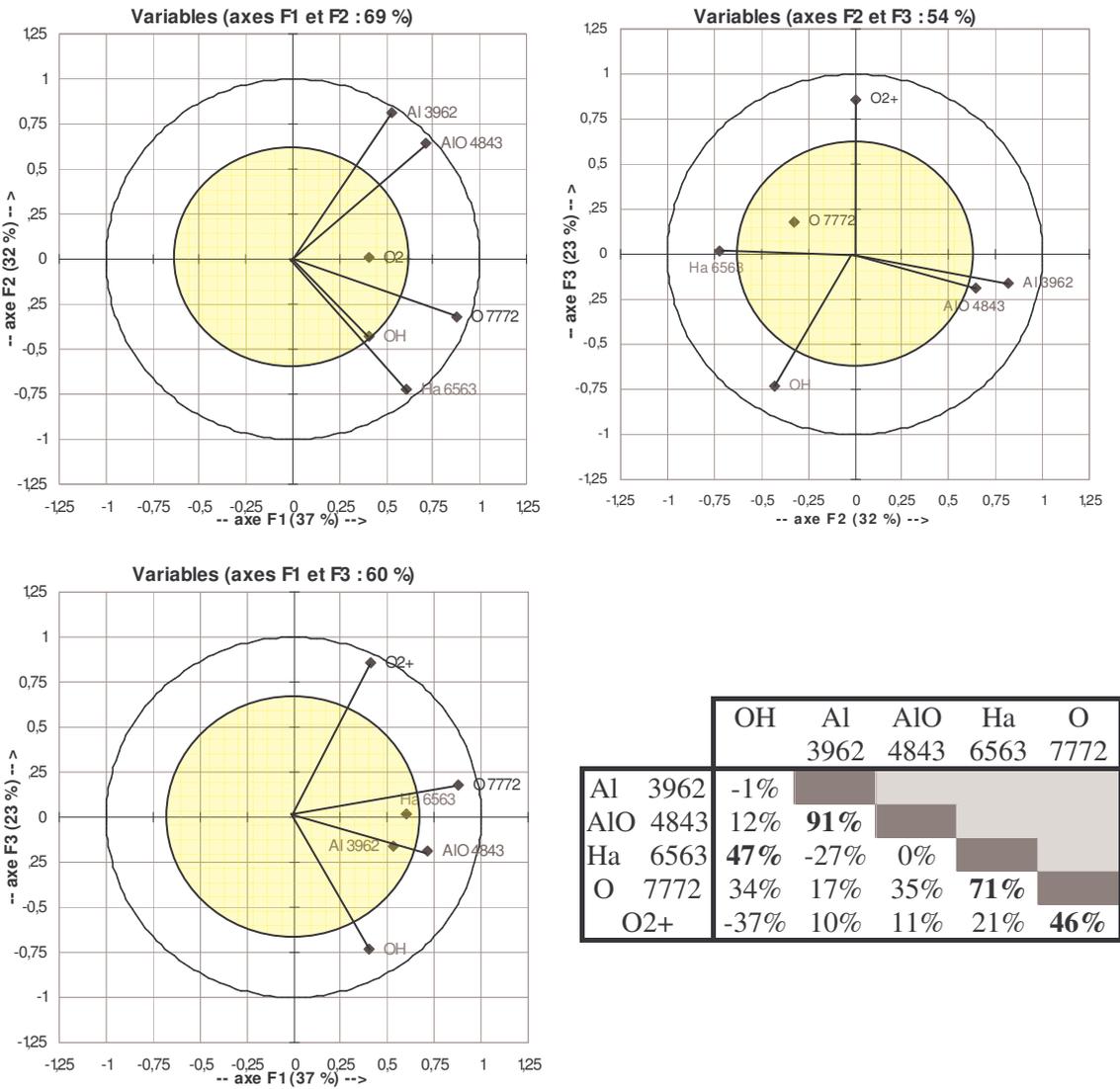


Figure C.24 Etude par ACP des relations entre les espèces du plasma

On constate qu'il existe une très forte corrélation entre les espèces du même type : d'une part l'aluminium atomique Al et l'aluminium oxydé sous forme AlO et d'autre part l'hydrogène atomique et la forme hydroxyle (OH). Une corrélation entre l'oxygène atomique et l'ion O_2^+ semble aussi exister mais elle est plus difficile à caractériser compte tenu des projections utilisées (seul le plan Axe1/Axe3 représente correctement ces deux variables).

De plus, ces graphiques révèlent une absence de corrélation entre l'aluminium atomique et les espèces oxygénées détectées dans le plasma (O atomique, O_2^+ , OH) ainsi qu'avec l'hydrogène atomique. L'hydrogène atomique est, en revanche, fortement corrélé avec l'oxygène atomique mais peu ou pas corrélé avec l'oxygène moléculaire.

I.5.2.d. Influence des propriétés des films sur les espèces oxygénées

(O, O₂⁺)

Les projections sur trois plans issus des trois premiers axes principaux de l'A.C.P. sont représentées sur la **Figure C.25** ainsi que le tableau des corrélations. La représentativité de ces plans est relativement faible (73% pour les trois plans) et peu de projections sont supérieures à 0,7. Dans ce cas, on complète l'analyse graphique par l'interprétation des coefficients de corrélations. Si on se reporte au tableau des corrélations, on constate que la prise de masse n'est pas influencée par la quantité d'oxygène atomique présent dans le milieu réactif mais qu'elle est relativement sensible à la présence d'O₂⁺. On observe le phénomène inverse pour la vitesse de dissolution.

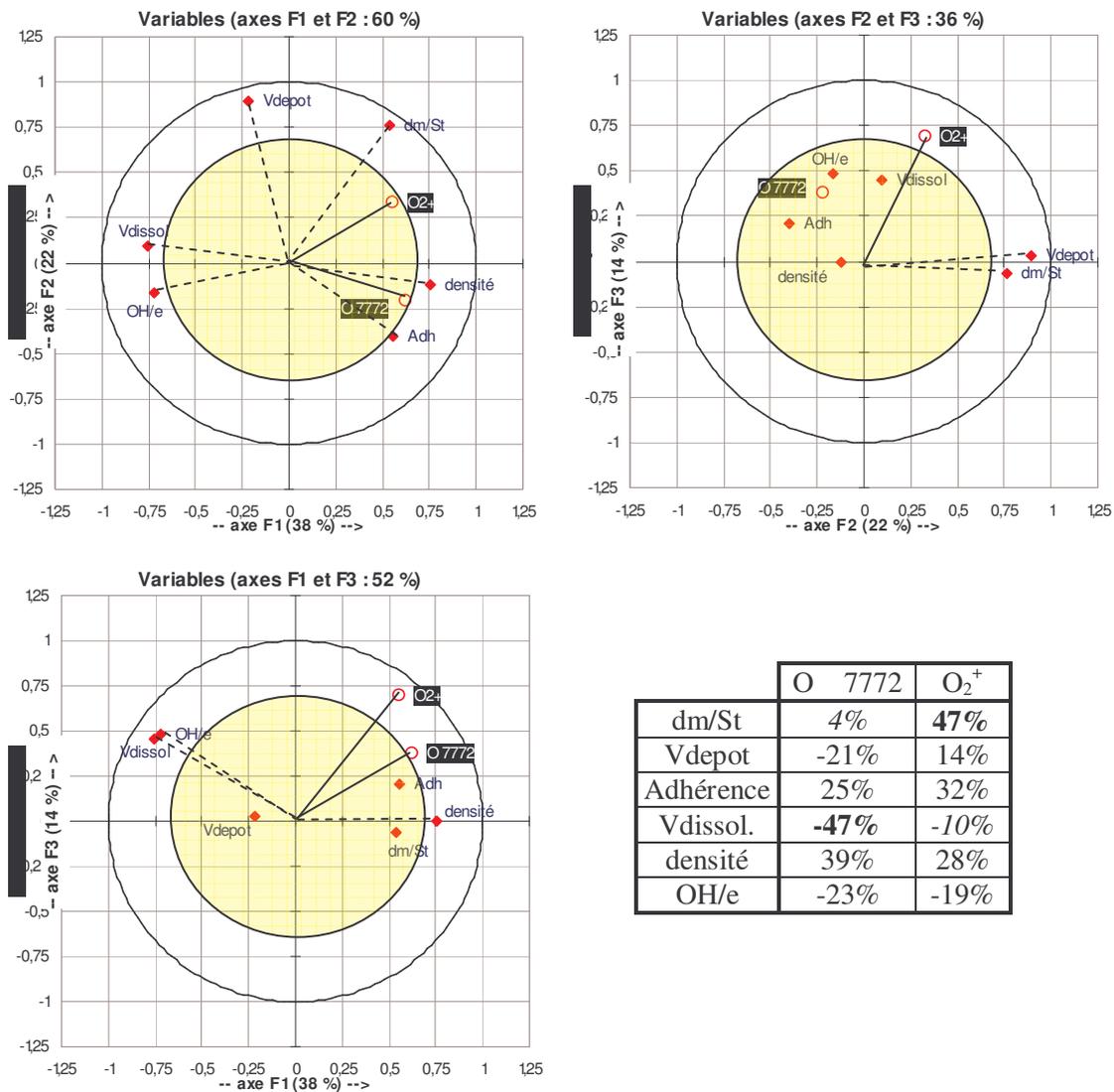


Figure C.25 Relations entre les espèces oxygénées du plasma et les propriétés des films

I.5.2.e. Influence des espèces contenant l'aluminium (Al, AIO)

La projection sur trois plans issus des trois premiers axes principaux de l'A.C.P. est représentée sur la **Figure C.26** ainsi que le tableau des corrélations. La représentativité de ces plans est correcte : on obtient 80% de l'information.

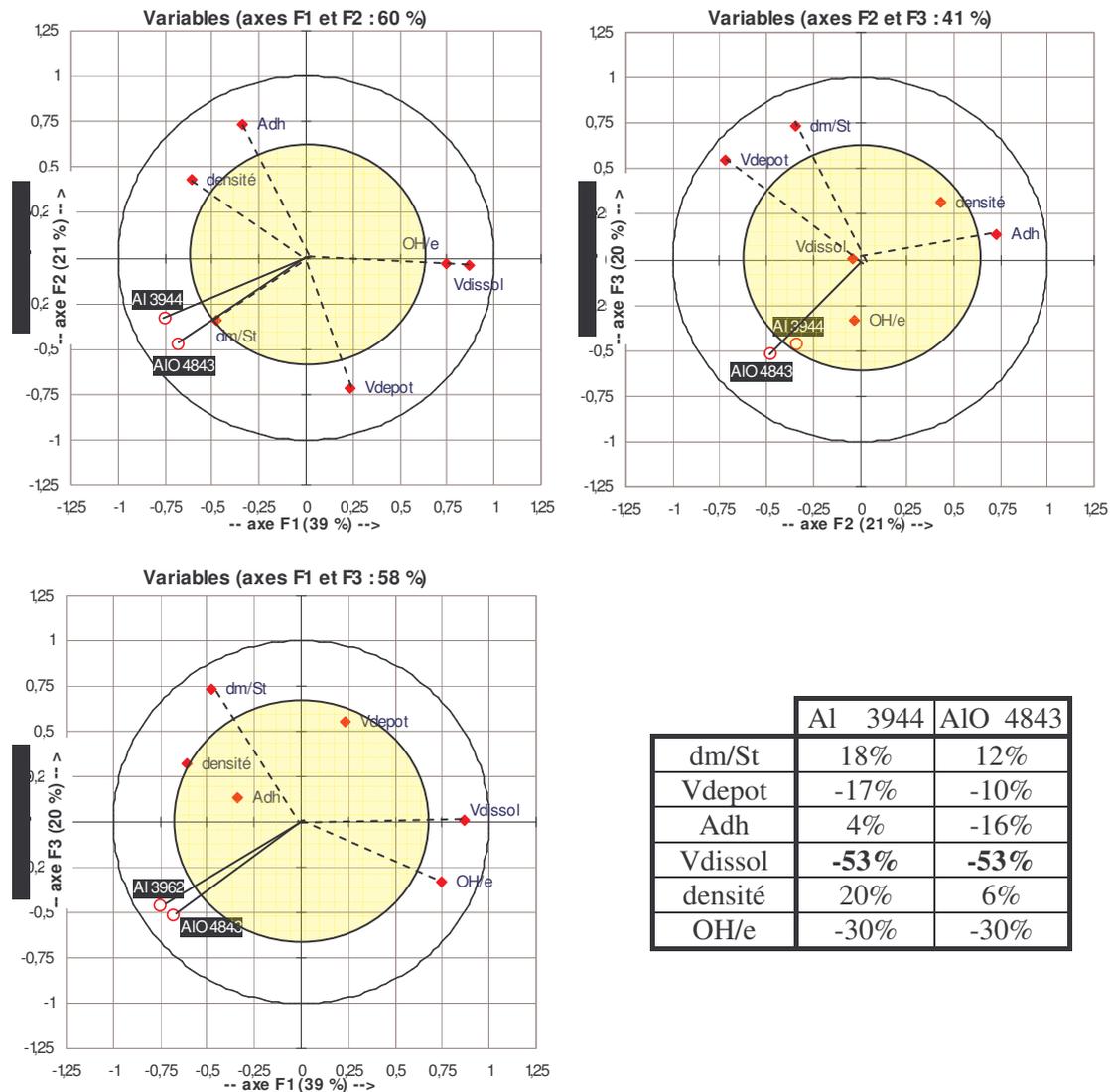


Figure C.26 Relations entre les espèces contenant de l'aluminium du plasma et les propriétés des films

Tout comme l'oxygène atomique, les espèces contenant de l'aluminium ne sont pas directement liées aux propriétés du dépôt mais, elles contrôlent, en partie, la qualité du film (Axe1/Axe2 : une quantité importante d'Al ou d'AIO entraîne une vitesse de dissolution plus faible).

On pouvait s'attendre à un résultat différent dans la mesure où les concentrations en Al et AIO et la prise de masse sont toutes les deux corrélées à la pression. Cependant, l'influence de ce dernier paramètre d'entrée dépend de la distance entre l'injecteur et le porte-substrat. La prise de masse

augmente si la pression diminue à une position donnée mais diminue globalement avec la distance. Les concentrations en Al et AlO, augmentent si la pression diminue et si la distance augmente.

I.5.2.f. Influence des espèces contenant l'hydrogène (H, OH)

Les projections sur trois plans issus des trois premiers axes principaux de l'A.C.P. sont représentées sur la **Figure C.27** ainsi que le tableau des corrélations. La représentativité de ces plans est limitée à 74% de l'information pour les trois axes.

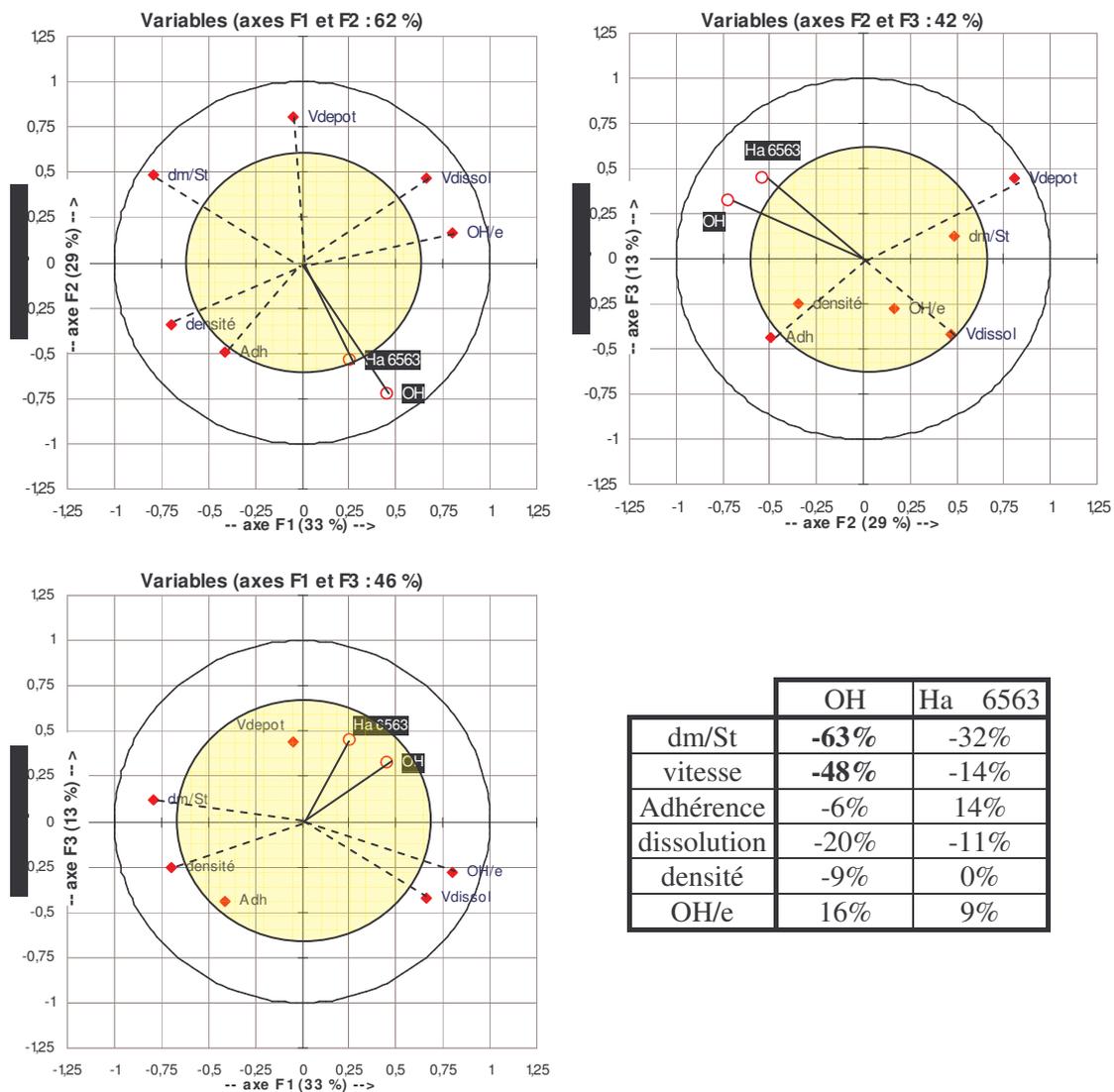


Figure C.27 Relations entre les espèces contenant de l'hydrogène du plasma et les propriétés des films

Les cercles de corrélations entre la composition en espèces hydrogénées (H, OH) et les propriétés des films ne permettent pas l'établissement de corrélation directe. Pour le groupement OH, il est cependant intéressant de noter la non-corrélation entre sa concentration en phase gazeuse et son

taux dans le film. En revanche, la corrélation est établie avec les cinétiques de dépôt. Donc, une prise de masse ou une vitesse de dépôt élevée se traduit par une concentration faible d'OH dans la phase gazeuse [1].

L'analyse de données a donc permis de dégager de nombreuses corrélations existantes tout d'abord entre les propriétés des films (réponses du système) et entre ces mêmes propriétés et celles du plasma.

I.6. CONCLUSION

A partir de 8 paramètres expérimentaux, un plan de criblage nous a permis dans un premier temps de classer les facteurs par ordre de leur importance sur les propriétés physicochimiques et les vitesses de dépôt des films. Ensuite, les paramètres les plus influents (pression dans l'enceinte, distance injecteur/substrat, température du substrat et polarisation radiofréquence) ont été optimisés à l'aide d'un plan à surfaces de réponses (réseau uniforme de **Doelhart**). De nombreuses analyses en composantes principales ont été réalisées par la suite, pour déterminer les corrélations existantes entre les propriétés du film ainsi que les corrélations existantes entre les propriétés des films et le plasma. Ce chapitre a donc mis en évidence la complémentarité et l'enchaînement de ces méthodes (voir **Figure C.28**). Nous avons donc mis en exemple dans ce chapitre la notion de construction d'un plan (**Plackett** et **Burmann**, **Doerlhart**), la description des analyses mathématiques d'un plan d'expériences (calcul des coefficients - méthode des moindres carrés), les notions d'analyses statistiques (méthode de **Lenth**, droite de **Henry**, Test de **Fisher**) ainsi que l'efficacité en terme visuel des analyses graphiques (diagramme en bâton, diagramme de Pareto, courbes isoréponses et surfaces de réponses).

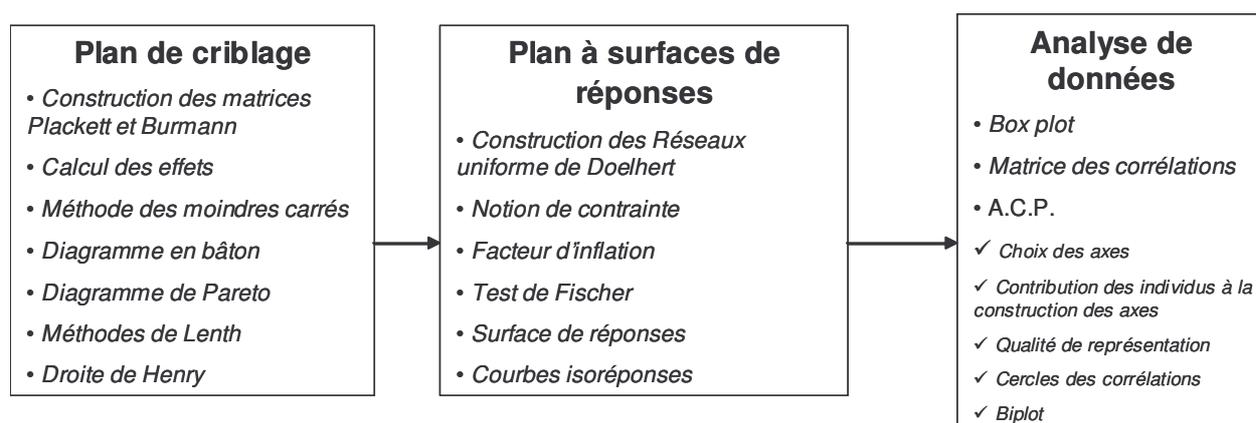


Figure C.28 Schéma récapitulatif des analyses du chapitre I

Chapitre II : Optimisation des couches minces de titanate de baryum et de strontium obtenues par ablation laser

Ces travaux ont été menés en collaboration avec **S. Liebus** qui a soutenu sa thèse [6] en novembre 2003 au sein de l'équipe Composants et Circuits Microélectroniques et Microoptiques (**C2M**) de l'Institut de Recherche en Communications Optiques et Microondes (**IRCOM**).

II.1. FORMALISATION DU PROBLEME

II.1.1. OBJECTIF

II.1.1.a. Présentation de l'étude

Le travail de recherche présenté ici a pour objectif l'optimisation des propriétés des films de titanate de baryum et de strontium, plus communément appelé BST, réalisés par ablation laser, par **S. Liebus**. Ce matériau possède des propriétés ferroélectriques permettant d'obtenir une variation de sa permittivité diélectrique sous l'effet d'un champ électrique statique ce qui en fait un excellent candidat pour des applications telles que les fonctions de filtrage accordable en fréquence.

Nous verrons dans cette partie, que toutes les analyses de données et les plans d'expériences développés, ont pour principal objectif l'optimisation de la valeur de la permittivité diélectrique. Nous allons tout d'abord présenter le dispositif expérimental utilisé et les principaux paramètres de dépôt pouvant influencer sur la valeur de la permittivité diélectrique.

II.1.1.b. Dispositif expérimental

Le dispositif expérimental d'ablation laser est schématisé sur la **Figure C.29**. Il est composé essentiellement d'un laser et d'une enceinte de dépôt sous vide reliée à un groupe de pompage. Les films de BST sont déposés sur des substrats d'oxyde de magnésium (MgO) et caractérisés après dépôt. Comme dans l'exemple précédent décrit dans le chapitre I, nous allons lister les variables d'entrée constituées par les différents paramètres de dépôt et les variables de sortie regroupant les résultats obtenus à partir de l'ensemble des outils de caractérisation.

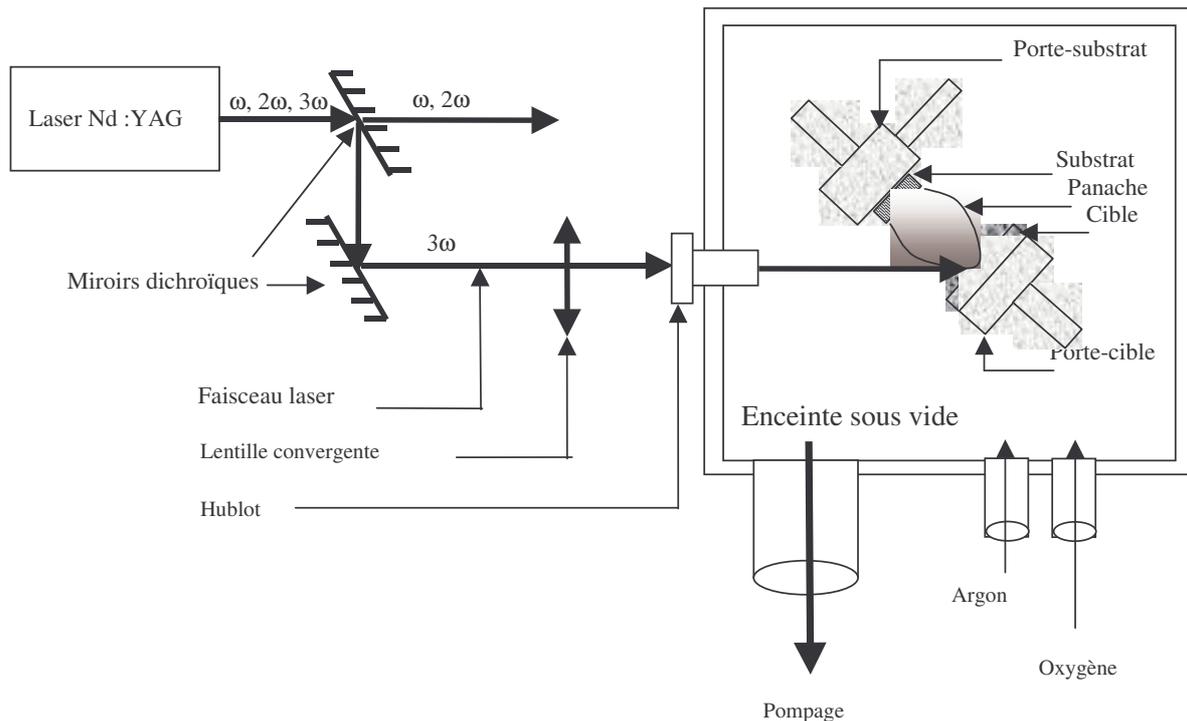


Figure C.29 Dispositif expérimental d'ablation laser

Les analyses de données et les plans d'expériences que nous allons présenter dans la suite de ce chapitre, ont été menés avec pour principaux objectifs :

- d'identifier les relations entre les variables d'entrée et de sortie,
- de corrélérer les propriétés des films aux facteurs d'entrée,
- de déterminer les conditions expérimentales conduisant à une mesure significative de la permittivité diélectrique. En effet, compte tenu de la sensibilité du dispositif utilisé et de la qualité des films de BST, il n'est pas toujours possible d'obtenir une mesure de la permittivité diélectrique ou parfois, cette mesure est inexploitable. Nous verrons comment, dans la suite de ce chapitre, en utilisant une méthode d'analyse de données multidimensionnelle (l'Analyse Factorielle Discriminante), il nous a été possible de prévoir les conditions expérimentales de dépôt conduisant à une mesure de la constante diélectrique,
- de maximiser la valeur de la constante diélectrique. L'expérimentateur cherche à déterminer l'évolution de la valeur de la permittivité en fonction des facteurs d'entrée permettant d'obtenir une forte valeur de permittivité.

Objectif : « Maximiser la valeur de la permittivité diélectrique »

II.1.2. REPONSES

Chaque dépôt est caractérisé par différents moyens et les valeurs des mesures réalisées ont permis de définir les variables de sortie ou réponses (**Tableau C.33**) suivantes :

- l'épaisseur e (μm) et la rugosité R_a (\AA) des couches mesurées à l'aide d'un profilomètre,
- l'organisation des phases cristallines présentes au sein du dépôt déterminée par analyse par diffraction des rayons X. Ce facteur de sortie, appelé **crystallisation**, prendra des valeurs comprises entre 0 et 2 en fonction de la qualité de la phase cristalline obtenue et sera évalué par l'expérimentateur,
- la permittivité diélectrique ϵ . Cette grandeur est mesurée par un dispositif de caractérisation hyperfréquence développé par l'équipe Circuits et Dispositifs Microondes (C.D.M.) de l'I.R.C.O.M. (voir **Chapitre III** de cette Partie).

Tableau C.33 Réponses et notations

Réponses	Notation	Unité
Epaisseur	e	μm
Rugosité	R_a	\AA
Permittivité diélectrique	ϵ	Sans unité
Cristallisation	Cr	Sans unité

II.1.3. FACTEURS

Nous avons choisi ici de présenter les variables d'entrée en décrivant les différentes phases intervenant avant, pendant et après le déroulement d'une expérimentation.

Ces différentes étapes sont :

- la fabrication des cibles de BST. Le matériau cible, de composition $\text{Sr}_x\text{Ba}_{1-x}\text{TiO}_3$ est réalisé au Laboratoire des Sciences des Procédés Céramiques et Traitement de Surface (**S.P.C.T.S.**). La **composition** souhaitée, à savoir : x mole (avec $0 < x < 1$) de strontium pour $(1 - x)$ mole de baryum et une mole de titane peut être ajustée. La valeur prise par x constitue un premier facteur,
- la mise sous tension de l'élément chauffant. Les substrats de MgO nettoyés ex-situ, sont préalablement placés sur un porte-substrat plaqué sur un élément chauffant. Le 2^{ème} facteur est donc la **température** des substrats pouvant atteindre $850\text{ }^\circ\text{C}$,

- la mise sous tension du laser. Le faisceau du laser est focalisé sur la surface de la cible au moyen d'une lentille convergente. L'énergie reçue par unité de surface de la cible est définie comme la **fluence** laser. Ce 3^{ème} facteur peut varier entre 1 et 5 J/cm²,
- la phase de dépôt. La **durée** du dépôt qui varie de quelques minutes à quelques heures selon l'épaisseur souhaitée et la **pression** régnant dans l'enceinte (domaine de variation : 0,05 – 0,3 Pa) constituent deux autres facteurs,
- enfin, après dépôt, un **recuit** ex-situ peut être réalisé. Ce facteur prendra la valeur 1 si les films de BST sont recuits et la valeur 0 dans le cas contraire.

Le **Tableau C.34** résume ces différents facteurs.

Tableau C.34 Facteurs et notations

Facteurs	Notation	Unité	Nature
Composition	x	Sans unité	Quantitatif
Température	T	°C	Quantitatif
Fluence	F	J/cm ²	Quantitatif
Durée du dépôt	t	min	Quantitatif
Recuit	R	Sans unité	Qualitatif

Dans un premier temps nous allons analyser les données existantes.

II.2. ANALYSE DE DONNEES EXISTANTES

Nous allons effectuer trois sortes d'analyses : une analyse en composantes principales pour identifier les relations entre les variables d'entrées et la réponse ϵ du système, une classification pour former des familles d'expériences et une analyse factorielle discriminante pour déterminer les conditions expérimentales conduisant à une mesure significative de la permittivité.

II.2.1. TABLEAU DE DONNEES INITIALES

La première chose à faire quand on récupère un tableau de données, c'est de le réorganiser suivant l'analyse que l'on souhaite effectuer. Le **Tableau C.36** et le **Tableau C.37** sont des sous ensembles du **Tableau C.35** qui représente le tableau initialement donné par l'expérimentateur.

Ces sous tableaux restructurés nous serviront de tableaux bruts pour les analyses à venir. A partir du **Tableau C.36**, nous ferons une **A.F.D.** et à partir du **Tableau C.37**, une **A.C.P.**

Tableau C.35 Ensemble des données

N° essai	Cibles (x)	Durée(mm)	Sustrat recuit	Pression	Fluence	Temperature	Cristallisation	Epsilon	Epais.(µm)	rugosité (A)
78	0,5	34	0 : pas recuit	0,075	3	850	0	350	2,2	660
57	0,3	60	0	0,075	3	810	1	165	1,92	1190
72	0,3	30	1 : recuit	0,075	3	850	2	310	1,9	960
54	0,3	60	0	0,075	3	795	1	270	1,76	1170
55	0,3	65	0	0,075	3	790	1	167	1,73	1300
58	0,3	60	0	0,075	3	805	1	251	1,7	770
86	0,8	35	0	0,075	3	850	2	325	1,654	790
59	0,3	60	0	0,075	3	820	1	174	1,53	870
87	0,5	30	0	0,075	3	850	1	290	1,511	860
56	0,3	60	0	0,075	3	800	1	204	1,51	1340
38	0,8	45	1	0,075	3	800	0	155	1,5	2400
85	0,8	30	0	0,075	3	850	2	360	1,36	680
89	0,8	26	0	0,075	3	850	2	450	0,909	550
75	0,3	6	0	0,075	3	850	2	300	0,869	660
88	0,8	23	0	0,075	3	850	2	450	0,8	420
51	0,3	30	0	0,075	2,5	790	1	373	0,7	640
69	0,3	6	0	0,075	3	850	2	500	0,58	740
76	0,3	8	0	0,075	3	850	2	730	0,37	610
73	0,3	20	0	0,075	3	850	1	erreur	1,78	980
45	0,5	26	0	0,075	3	800	0	erreur	1,157	900
47	0,5	29	0	0,075	3	800	0	erreur	0,97	1200
37	0,8	30	0	0,075	3	800	2	erreur	0,87	290
48	0,5	27	0	0,075	3	800	1	erreur	0,8	870
46	0,5	23	0	0,075	3	800	2	erreur	0,7	950
43	0,5	21	0	0,075	3	800	1	erreur	0,63	950
83	0,8	16	0	0,075	3	850	1	erreur	0,61	690
52	0,3	20	0	0,075	2,5	800	0	erreur	0,56	1730
71	0,3	8	0	0,075	3	850	1	erreur	0,48	780
53	0,3	18	0	0,075	2,5	820	1	erreur	0,48	690
42	0,5	12	0	0,075	3	800	1,5	erreur	0,37	1300
68	0,3	4	0	0,075	3	850	2	erreur	0,34	580
44	0,5	6	0	0,075	3	800	2	erreur	0,28	860
70	0,3	1	0	0,075	3	850	0		0,2	440
91	0,5	1	1	0,075	3	850		erreur	0,04	410
11	0,8	15	0	0,1	3	500	0	pas mesura	0,42	472
12	0,8	15	0	0,1	3	600	0	pas mesura	0,44	264
13	0,8	15	0	0,1	3	700	0	pas mesura	0,33	133
14	0,8	15	0	0,1	3	800	0	pas mesura	0,81	370
15	0,8	30	0	0,1	4	400	0	pas mesura		
16	0,8	30	0	0,1	4	500	0	pas mesura		
17	0,8	30	0	0,1	4	650	0,5	pas mesura		
18	0,8	30	0	0,1	4	800	0,5	pas mesura		
49	0,3	15	0	0,075	2,5	790	1	pas mesura		
20	0,8	30	1	0,1	3	850	1,5	pas mesura		
21	0,8	30	1	0,1	3	800	0	pas mesura		
22	0,8	30	1	0,1	3	800	0	pas mesura		
27	0,8	30	1	0,1	3	850	0	pas mesura		
65	0,3	15	0	0,075	3	790	2	pas mesura		
66	0,3	15	0	0,075	3	850	2	pas mesura		
74	0,3	12	0	0,075	3	850	2			
10	0,8	15	0	0,1	3	400	0	pas mesura	1,1	400
32	0,8	30	0	0,05	3	850	1	pas mesura	0,8	300
33	0,8	30	0	0,075	3	850	1,5	pas mesura	1,4	640
34	0,8	30	0	0,075	3	825	2	pas mesura	0,8	580
35	0,8	30	1	0,075	3	825	2	pas mesura	0,7	490
36	0,8	30	0	0,075	3	800	2	pas mesura	0,7	870
19	0,8	30	0	0,1	3	850	1	pas mesura	0,24	2500
50	0,3	15	0	0,075	2,5	790	1,5	pas mesura	0,56	1030
62	0,3	15	0	0,1	3	850	2	pas mesura	0,59	770
63	0,3	17	0	0,075	3	850	2	pas mesura	0,79	420
64	0,3	15	0	0,075	3	820	1,5	pas mesura	0,65	830
31	0,8	30	0	0,1	3	850	1	pas mesura	0,5	520
67	0,3	2	0	0,075	3	850	1,5		0,39	470
28	0,8	30	1	0,05	3	850	1,5	pas mesura	0,48	820
29	0,8	30	0	0,3	3	850	0	pas mesura	0,4	330
77	0,5	45	0	0,075	3	500		pas mesura	2,06	1220
79	0,8	30	0	0,075	3	500		pas mesura	0,82	370
80	0,8	15	0	0,075	3	500		pas mesura	0,47	260
84	0,8	20	0	0,075	3	850	2		0,46	360
90	0,5	2	1	0,075	3	850	2		0,12	440
30	0,8	30	0	0,1	3	850	0	pas mesura	0,35	230

Tableau C.36 Tableau de données brutes de l'A.F.D.

Number	x	t	T	ε
78	0,5	34	850	A : mesurable
57	0,3	60	810	A
72	0,3	30	850	A
54	0,3	60	795	A
55	0,3	65	790	A
58	0,3	60	805	A
86	0,8	35	850	A
59	0,3	60	820	A
87	0,5	30	850	A
56	0,3	60	800	A
38	0,8	45	800	A
85	0,8	30	850	A
89	0,8	26	850	A
75	0,3	6	850	A
88	0,8	23	850	A
51	0,3	30	790	A
69	0,3	6	850	A
76	0,3	8	850	A
73	0,3	20	850	B : erreur
45	0,5	26	800	B
47	0,5	29	800	B
37	0,8	30	800	B
48	0,5	27	800	B
46	0,5	23	800	B
43	0,5	21	800	B
83	0,8	16	850	B
52	0,3	20	800	B
71	0,3	8	850	B
53	0,3	18	820	B
42	0,5	12	800	B
68	0,3	4	850	B
44	0,5	6	800	B
70	0,3	1	850	B
84	0,8	20	850	B
67	0,3	2	850	B
11	0,8	15	500	C : non mesurable
12	0,8	15	600	C
13	0,8	15	700	C
14	0,8	15	800	C
10	0,8	15	400	C
32	0,8	30	850	C
33	0,8	30	850	C
34	0,8	30	825	C
35	0,8	30	825	C
36	0,8	30	800	C
19	0,8	30	850	C
50	0,3	15	790	C
62	0,3	15	850	C
63	0,3	17	850	C
64	0,3	15	820	C
31	0,8	30	850	C
28	0,8	30	850	C
29	0,8	30	850	C
90	0,5	2	850	C
30	0,8	30	850	C

Tableau C.37 Tableau de données brutes de l'A.C.P.

N°essai	x	t	T	ϵ	e
78	0,5	34	850	350	2,2
57	0,3	60	810	165	1,92
72	0,3	30	850	310	1,9
54	0,3	60	795	270	1,76
55	0,3	65	790	167	1,73
58	0,3	60	805	251	1,7
86	0,8	35	850	325	1,654
59	0,3	60	820	174	1,53
87	0,5	30	850	290	1,511
56	0,3	60	800	204	1,51
38	0,8	45	800	155	1,5
85	0,8	30	850	360	1,36
89	0,8	26	850	450	0,909
75	0,3	6	850	300	0,869
88	0,8	23	850	450	0,8
51	0,3	30	790	373	0,7
69	0,3	6	850	500	0,58
76	0,3	8	850	730	0,37

Nous pouvons déjà faire quelques remarques en jetant un premier regard sur le **Tableau C.35**.

- on s'aperçoit qu'il y a des cases vides, sans valeurs (nous les appelons aussi des **trous**). Notons, qu'il est souhaitable pour faire de l'analyse de données de ne pas avoir de données manquantes dans un tableau (voir **Partie B**). Il faudra donc réorganiser le tableau et enlever dans premier temps les lignes sans données complètes.
- il y a des variables quantitatives qui varient très peu comme la pression qui prend deux valeurs (0,75 et 1) ainsi que la fluence qui en prend trois (2,5 ; 3 et 4). Ces variables ne fluctuent pas assez pour être analysées (voir **Partie B**), elles ne seront donc pas analysées.

Donc, à partir du **Tableau C.35**, si on enlève toutes les lignes contenant des données manquantes, les colonnes Pression, fluence, substrat recuit et les colonnes (cristallisation, épaisseur, rugosité) correspondant aux réponses que l'on ne souhaite pas momentanément analyser, on obtient le **Tableau C.36**. Le **Tableau C.37** regroupe tous les essais du **Tableau C.36** pour lesquels on obtient une réponse mesurable de la permittivité, auxquels on a rajouté une colonne correspondant à la valeur des épaisseurs mesurées.

II.2.2. PRE-TRAITEMENT DES DONNEES

II.2.2.a. **Box plot**

La **Figure C.30** représente les boîtes à moustaches des variables : x, t, T, e et ϵ issues du tableau de données brutes de l'A.C.P.

On remarque que :

- L'individu **76** est une expérience atypique, il possède une valeur très élevée de la permittivité diélectrique (730),
- x ne prend que trois valeurs (la boîte ne présente pas de pattes). L'expérimentateur n'a testé que trois valeurs de x pour des contraintes de coût et de procédé difficile à mettre en œuvre.

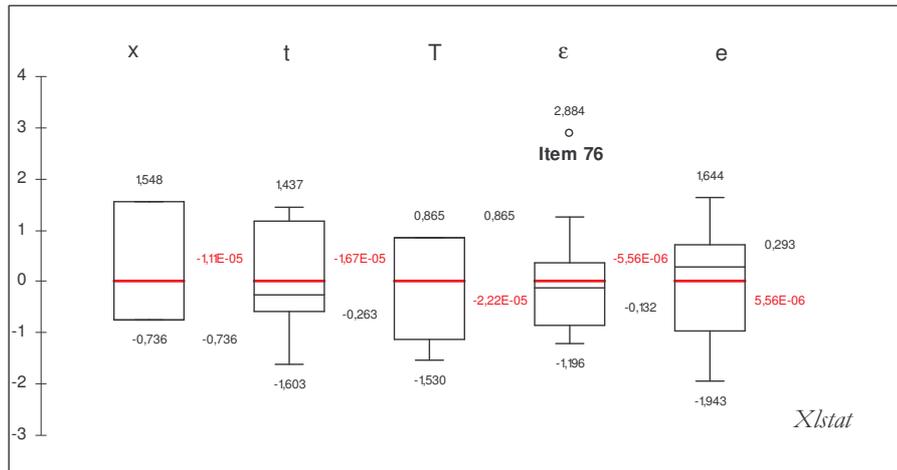


Figure C.30 Box plot des variables du **Tableau C.37**

II.2.2.b. Matrice des corrélations et matrice XY

L'analyse bidimensionnelle détermine les relations des variables deux à deux. La matrice des corrélations répertorie toutes les valeurs des coefficients de corrélations entre les cinq variables (**Tableau C.38**) et la matrice XY est une représentation graphique (**Figure C.31**) de ces corrélations.

Tableau C.38 Matrice des corrélations

	Cibles (x)	Durée(mm)	Temperature	Epsilon	Epais.(μ m)
Cibles (x)	1	-0,289	0,437	0,330	-0,151
Durée(mm)	-0,289	1	-0,754	-0,793	0,654
Temperature	0,437	-0,754	1	0,643	-0,195
Epsilon	0,330	-0,793	0,643	1	-0,627
Epais.(μ m)	-0,151	0,654	-0,195	-0,627	1

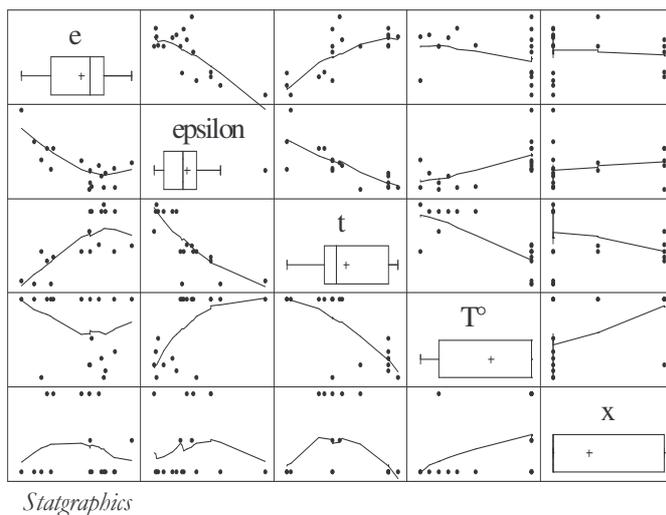


Figure C.31 Multiple XY

Nous remarquons que le **Tableau C.38** contient des coefficients de corrélation significatifs (en gras dans le tableau). Ces représentations bidimensionnelles des corrélations vont appuyer les résultats de l'**A.C.P.**. En effet, en cas de doute sur des corrélations, on calculera le coefficient de détermination pour estimer la *force de liaison* (**Partie B**).

Pour avoir une vue globale et plus lisible de ces corrélations, nous procédons maintenant à une analyse multidimensionnelle des données : l'**A.C.P.**

II.2.3. ANALYSE EN COMPOSANTES PRINCIPALES

II.2.3.a. Bilan des analyses précédentes

La boîte à moustache de la variable permittivité (**Figure C.30**) nous indique qu'il existe une expérience atypique (individu **76**). Pour ne pas fausser l'analyse, nous avons décidé de mettre cet individu en tant qu'individu supplémentaire dans l'analyse en composantes principales.

II.2.3.b. Part d'information portée par les axes

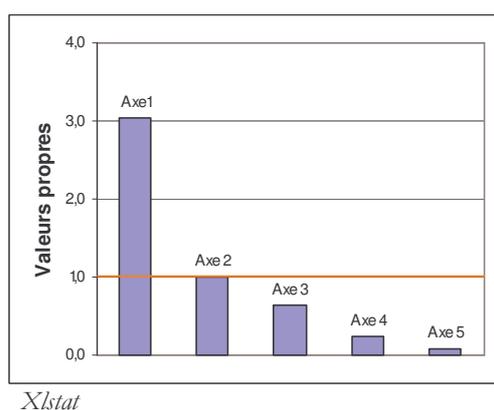


Figure C.32 Histogramme des valeurs propres

L'histogramme des valeurs propres (**Figure C.32**), nous indique qu'il suffit de regarder les résultats de l'A.C.P. sur deux axes (axes **1** et **2**). On rappelle que la part de variation expliquée par ces axes (valeur propre) doit être supérieure ou égale à celle de la variable initiale (égale à 1). Or, la valeur propre de l'axe **1** est nettement supérieure à 1 (3,048) et celle de l'axe **2** est proche de 1 (0,9930).

On remarque tout de même que le bâton représentant la valeur propre de l'axe 1 dans l'histogramme est nettement plus grand que les autres, ce qui traduit que l'information contenue dans le tableau brut initial (**Tableau C.37**) sera essentiellement représentée par l'axe 1.

Le pourcentage de variance et la variance cumulée pour chaque axe confirmeront cette hypothèse (**Tableau C.39**).

Tableau C.39 Part d'information donnée par les axes

	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
Valeur propre	3,048	0,993	0,643	0,231	0,085
% variance	60,959	19,869	12,856	4,624	1,691
% cumulé	60,959	80,828	93,684	98,309	100,000

Les axes 1 et 2 expliquent à eux seuls les 4/5 (80%) de l'information contenue dans le tableau de données brutes (**Tableau C.37**), dont plus de 60 % est attribué à l'axe 1.

Nous allons à présent estimer la contribution des individus à la construction des axes 1 et 2.

Plan principal choisi : Axe 1 / Axe 2

II.2.3.c. Contribution des individus au calcul des axes

D'après les histogrammes des contributions (**Figure C.33**), les individus 69, 88, 89 et 55 contribuent fortement à la construction de l'axe 1 et les individus 51, 69 et 86 à la construction de l'axe 2.

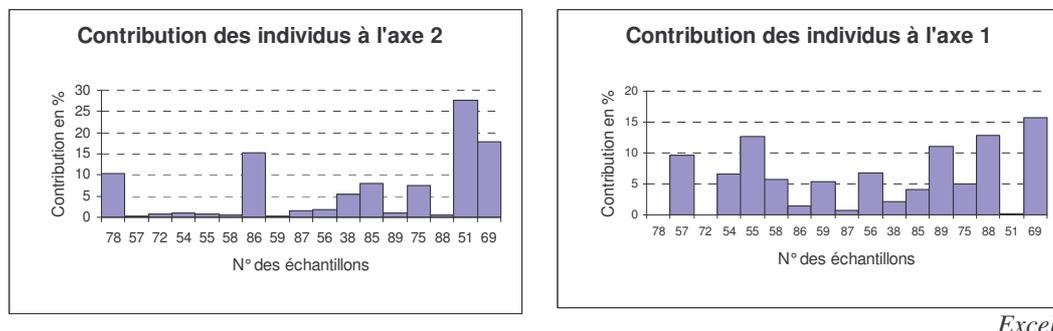
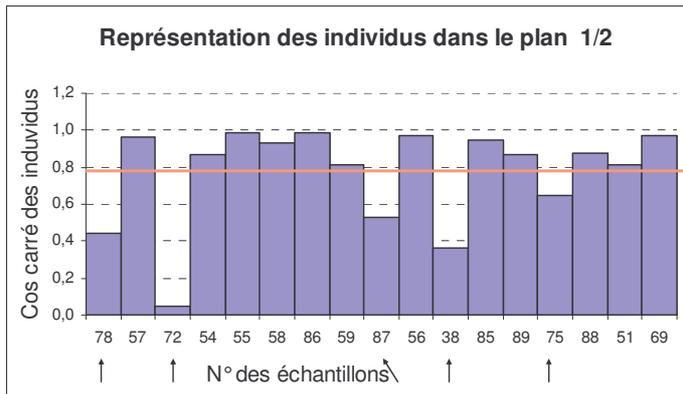


Figure C.33 Contribution des individus au calcul des axes 1 et 2

Dans l'ensemble, on peut dire que l'axe 1 est dû à un ensemble assez homogène d'individus ; cette axe sera considéré comme « stable » et nous pourrons lui accorder une signification aux représentations graphiques. L'axe 2 est moins stable que l'axe 1, puisque l'individu 51 contribue fortement à sa construction mais il n'explique que 20% de l'information du tableau de données. Par conséquent, il ne va pas influencer l'axe 1 qui a lui seul représente plus de 60% de l'information du **Tableau C.37**. Donc, nous pourrons considérer l'ensemble des deux axes comme stables.

II.2.3.d. Qualité de représentation des individus dans le plan principal

L'histogramme de la **Figure C.34** représente la somme des cosinus carré sur les axes 1 et 2.

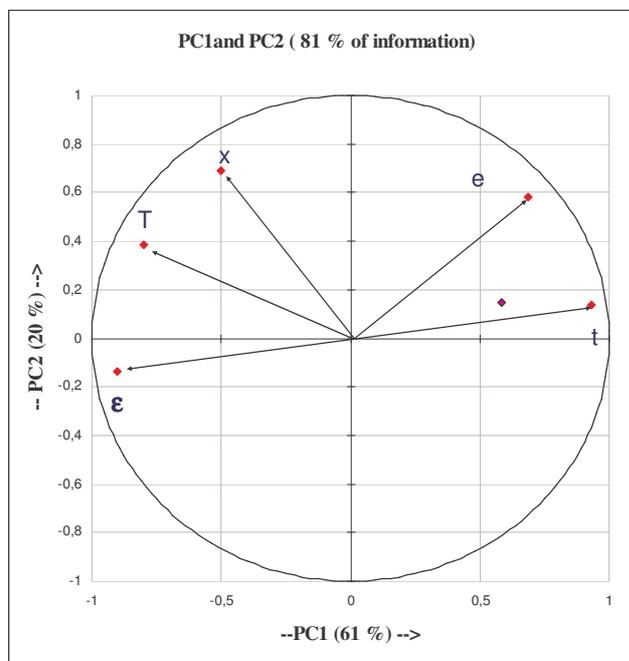


Excel

Figure C.34 Qualité de représentation des individus dans le plan principal

II.2.3.e. Analyse graphique

■ Cercle des corrélations



Xlstat

Rappelons ici que nous cherchons à relier la réponse ϵ avec les quatre autres facteurs (T , x , e , t) mais aussi, à corrélérer les variables d'entrée en elles.

Figure C.35

Cercle des corrélations

Le cercle des corrélations **Figure C.35** nous indique que :

- l'épaisseur e est corrélée positivement avec la durée de dépôt t (ce qui est logique pour l'expérimentateur puisque la vitesse de dépôt correspond à une épaisseur déposée par unité de temps) [6],
- x est quasi indépendant de l'épaisseur e et donc de la durée t ,
- la température T est inversement corrélée à la durée de dépôt t (ce qui n'est pas une conclusion logique pour l'expérimentateur!). La température T est également quasi

indépendante de l'épaisseur e . Nous interpréterons plus précisément ces résultats dans la suite de ce paragraphe,

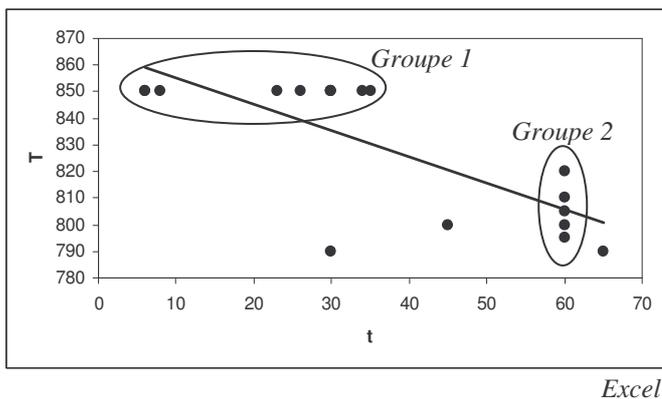
- ϵ est corrélée négativement avec l'épaisseur e et la durée t du dépôt et positivement avec la température T .

En conclusion, si l'on souhaite augmenter la valeur de la permittivité diélectrique, il faut déposer peu de temps à forte température, ce qui entraînera une couche peu épaisse. Ces résultats sont en accord avec les travaux menés précédemment au laboratoire par **S. Liebus [6]**.

$\epsilon \nearrow$: dépôt peu épais, courte durée de dépôt et forte température.

Attention !!!!! Toutes ces conclusions ne sont valables que pour les données figurant dans le tableau brut (**Tableau C.37**).

Regardons par exemple, la conclusion sur la corrélation négative entre la température T et la durée du dépôt t . L'expérimentateur sait pertinemment que ces deux facteurs sont indépendants.



Ainsi, la conclusion donnée par le cercle des corrélations peut paraître fausse pour l'expérimentateur, et c'est souvent dans ces circonstances que ces méthodes peuvent être rejetées !

La température T et le temps de dépôt t étant des paramètres connus comme indépendants par l'expérimentateur, pourquoi obtient-on une information que l'on peut dire « faussée » ?

Figure C.36 Nuage de points de T en fonction de t

Si on trace le nuage de points (**Figure C.36**) de T en fonction de t des expériences du tableau brut, on observe que l'expérimentateur, sans le savoir, a corrélé négativement ses réglages de la température et de la durée de dépôt. En effet, en appliquant la démarche « essai-erreur » (voir **Partie A**) qui consiste à faire varier un seul paramètre à la fois, l'expérimentateur a opté pour une variation des paramètres T et t de façon dépendante.

Ainsi, nous pouvons observer deux groupes d'expériences dans la **Figure C.36** : le groupe 1 consiste à fixer T (850°C) et à faire varier t ; dans le **groupe 2**, t est fixé (60 min) et T varie. Nous pouvons aussi noter que seulement trois individus se détachent des deux groupes.

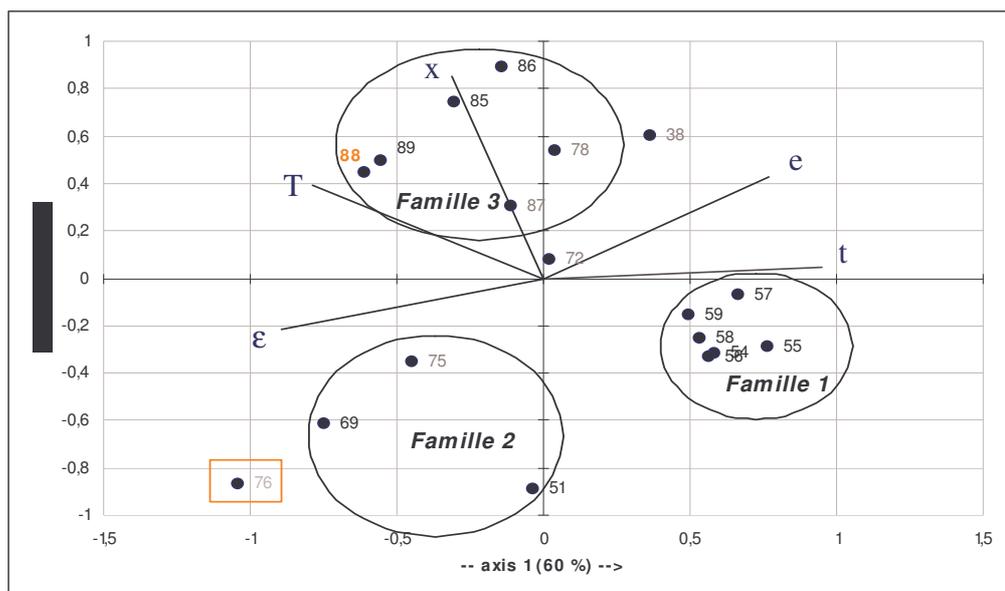
Il faut donc se méfier des conclusions hâtives que l'on peut déduire d'une analyse de données car, dire que les variables \mathbf{T} et \mathbf{t} sont corrélées sur l'ensemble du domaine expérimental délimité par les valeurs minimales et maximales du tableau de données peut s'avérer faux. Ici, on peut juste conclure sur les expériences du tableau et non sur l'ensemble du domaine expérimental.

▪ Biplot

Nous rappelons que le biplot (**Figure C.37**) est une représentation graphique où les variables et les individus sont simultanément représentés. Les individus supplémentaires peuvent être représentés sur ce graphique. Ainsi, nous pouvons voir de façon distincte que l'expérience **76** est éloignée des autres expériences. De plus, nous pouvons former des familles d'expériences :

- *Famille 1* : cette famille regroupe les expériences réalisées à basse température \mathbf{T} , avec une longue durée de dépôt \mathbf{t} et un faible pourcentage de strontium \mathbf{x} . Les individus de la *famille 1* ont une faible valeur de la permittivité ;
- *Famille 2* : cette famille regroupe les expériences réalisées avec un faible pourcentage de strontium \mathbf{x} et une courte durée de dépôt \mathbf{t} , donc une faible épaisseur \mathbf{e} . Les individus de la *famille 2* conduisent à une forte valeur de la permittivité ;
- *Famille 3* : cette famille regroupe les expériences réalisées avec une température \mathbf{T} élevée.

De plus, si on projette l'individu **88** sur les axes des variables, on remarque conformément aux conclusions établies dans le paragraphe précédent que cette expérience a forte valeur de $\mathbf{\epsilon}$, a bien été obtenue avec une faible épaisseur \mathbf{e} , une courte durée de dépôt \mathbf{t} et une forte température \mathbf{T} .



Xlstat

Figure C.37 Biplot dans le plan principal

Note : les individus grisés sont de mauvaise qualité de représentation dans le plan principal donc, nous ne pouvons tirer aucune conclusion sur la projection de ces expériences dans le biplot.

Une autre analyse appelée **Classification Ascendante Hiérarchique** (voir **Partie B**), complète de manière plus précise le biplot et permet d'observer les différentes familles d'individus.

II.2.4. CLASSIFICATION ASCENDANTE HIERACHIQUE

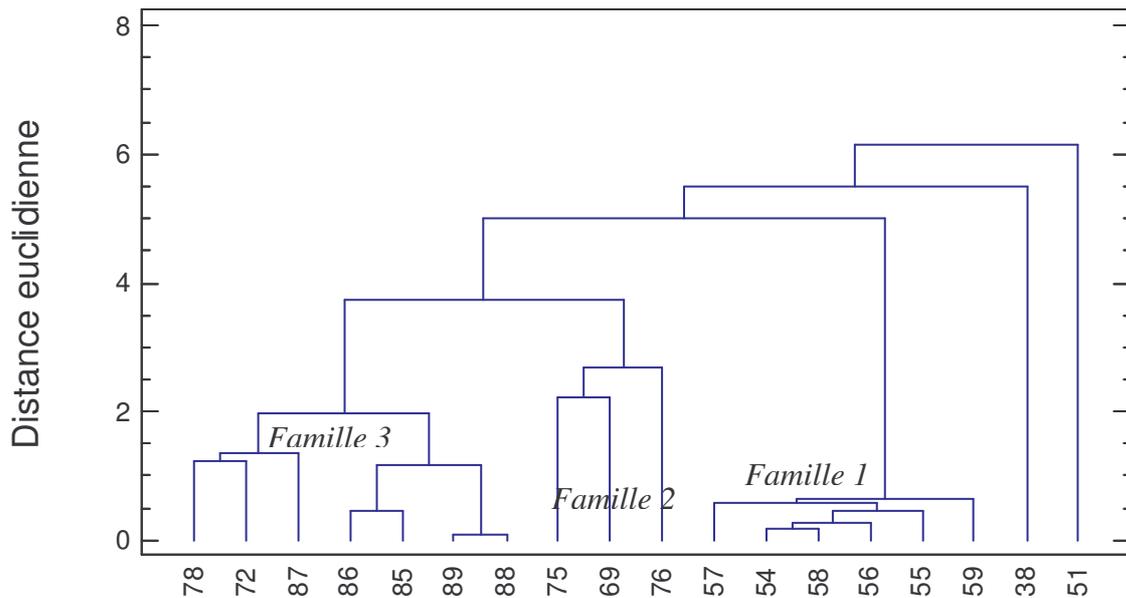


Figure C.38 Dendrogramme

En observant le dendrogramme de la **Figure C.38**, on retrouve les familles d'expériences précédemment citées. Cette représentation complète l'analyse graphique de l'A.C.P. (Biplot) car elle fournit des informations supplémentaires et plus précises sur les familles d'expériences. Par exemple, l'individu 51 que nous avons classé dans la famille 2 du Biplot apparaît ici comme un individu solitaire.

II.2.5. CONCLUSION SUR L'ANALYSE DE DONNEES

L'**A.C.P.** nous a permis de dégager des conclusions sur les influences des facteurs d'entrée (durée de dépôt, composition de la cible, température) et de regrouper les expériences en familles mais, elle ne permet pas d'obtenir les réglages exacts de ces paramètres pour optimiser la réponse ϵ . Ainsi, il nous a semblé intéressant de mettre en œuvre un plan d'optimisation.

II.3. OPTIMISATION

II.3.1. OBJECTIFS DE L'OPTIMISATION

Comme nous l'avons dit précédemment, l'objectif ici est de trouver des réglages plus précis des facteurs T , t et x de manière à maximiser la réponse ϵ du système.

II.3.2. FACTEURS -NIVEAUX-DOMAINES EXPERIMENTAL

Nous avons choisi d'optimiser le système à l'aide de trois facteurs ($k = 3$) : la température, la durée de dépôt et la composition de la cible. Nous avons choisi un **domaine** expérimental **sphérique** pour nous intéresser à ce qui se passe autour du réglage de référence. La matrice choisie est une **Doehlert**, précédemment détaillée dans le **chapitre 1** de cette Partie.

Les valeurs extrêmes du domaine expérimental sont représentées dans le **Tableau C.40**.

Tableau C.40 Domaine expérimental

	Facteur	Unité	Centre	Umin	Umax
U1	Température	°C	800	750	850
U2	Durée	min	37,5	15	60
U3	x		0,3	0,2	0,4

Au cours d'un brainstorming, nous avons dégagé les facteurs principaux du plan d'expériences ainsi que la difficulté de modification des niveaux de ces facteurs. Nous avons annoté ces difficultés à l'aide de symboles sur le diagramme d'**Ishikawa** (**Figure C.39** et **Figure C.40**).

Symbole	Groupe	Difficulté de modification des niveaux des facteurs
○	1	Difficile
◉	2	Assez difficile
●	3	Assez facile
●	4	Facile

Figure C.39 Difficulté de modification des niveaux des facteurs

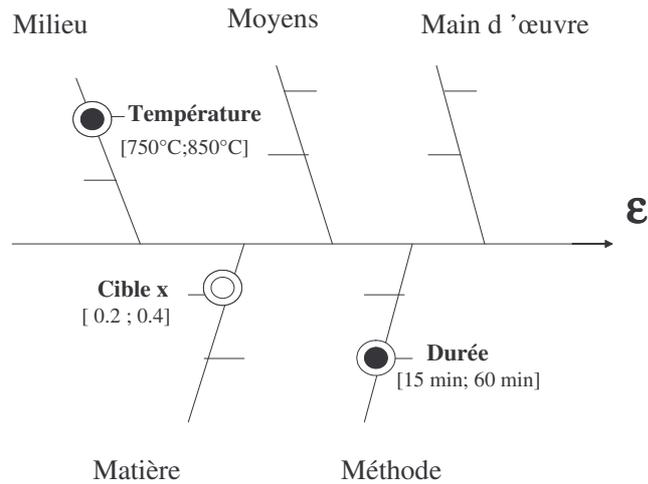


Figure C.40 Diagramme d'Ishikawa

II.3.3. MODELE EMPIRIQUE

Dans le cas présent, le modèle polynomial est de forme générale:

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{11} x_1^2 + b_{22} x_2^2 + b_{33} x_3^2 + b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{23} x_2 x_3 + \text{résidus}$$

Rappelons que le nombre d'inconnues d'un polynôme de degré 2 est établi à partir de la formule suivante :

$$p = \frac{(k+2)!}{k!2!} \Rightarrow p = \frac{(k+2)(k+1)}{2} \text{ on obtient donc : } p = 10$$

II.3.4. CONSTRUCTION DU PLAN D'EXPERIENCES

Les matrices de **Doehlert** sont des matrices asymétriques pour lesquelles tous les facteurs ne possèdent pas le même nombre de niveaux et dont le nombre de traitements expérimentaux distincts obéit à la relation suivante :

$$N = k^2 + k + 1$$

soit

$$N = 3^2 + 3 + 1 = 13 \text{ expériences à réaliser.}$$

La matrice d'expériences (**Tableau C.41**) associée à un réseau uniforme de **Doehlert** pour trois facteurs, nous donne le plan d'expérimentations à réaliser (**Tableau C.42**).

Tableau C.41 Matrice d'expériences

N°Exp	X1	X2	X3
1	1	0	0
2	-1	0	0
3	0,5	0,866	0
4	-0,5	-0,866	0
5	0,5	-0,866	0
6	-0,5	0,866	0
7	0,5	0,2887	0,8165
8	-0,5	-0,2887	-0,8165
9	0,5	-0,2887	-0,8165
10	0	0,5774	-0,8165
11	-0,5	0,2887	0,8165
12	0	-0,5774	0,8165
13	0	0	0

Les facteurs X_1 , X_2 et X_3 prennent respectivement 5, 7 et 3 niveaux. Le facteur x étant difficile à faire varier (voir **Figure C.40**), il lui sera attribué seulement 3 niveaux. Les autres paramètres étant faciles à faire varier, nous avons choisi 5 niveaux pour la température T et 7 niveaux pour la durée de dépôt t . On obtient donc :

X_1 : Température T – 5 niveaux
 X_2 : Durée t – 7 niveaux
 X_3 : composition x – 3 niveaux

Tableau C.42 Plan d'expérimentations

N°Exp	Température	Durée	x
	°C	min	
E1	850	37,5	0,3
E2	750	37,5	0,3
E3	825	60	0,3
E4	775	15	0,3
E5	825	15	0,3
E6	775	60	0,3
E7	825	45	0,4
E8	775	30	0,2
E9	825	30	0,2
E10	800	52,5	0,2
E11	775	45	0,4
E12	800	22,5	0,4
E13	800	37,5	0,3

Le **Tableau C.43** récapitule les objectifs du plan à surface de réponses.

Tableau C.43 Récapitulatif de l'étude du plan à surface de réponses

Objectif de l'étude	Etude dans un domaine expérimental: Surface de Réponses
Nombre de variables	3
Nombre d'expériences	13
Nombre de coefficients	10
Nombre de réponses	1

Avant de réaliser ces expériences, à partir du tableau de données initiales, on remarque que la mesure de la permittivité diélectrique du dépôt n'est pas toujours réalisable. Il est donc souhaitable avant de faire le plan de savoir si nous allons obtenir une valeur de la réponse. Pour cela, on utilise une méthode d'analyse de données (**Partie B**) appelée l'**Analyse Factorielle Discriminante (A.F.D.)**.

« Il faut s'assurer avant la réalisation des essais, que le plan d'expérimentation est capable d'atteindre les objectifs de l'expérience avec la précision souhaitée »

II.4. ANALYSE FACTORIELLE DISCRIMINANTE

Nous rappelons que l'**A.F.D.** s'intéresse aux relations entre une variable qualitative et un ensemble de variables quantitatives : il s'agit de relier l'appartenance d'un individu à une classe donnée aux valeurs qu'il prend pour les valeurs quantitatives.

La mesure de la permittivité diélectrique du dépôt n'est pas toujours réalisable, nous obtenons trois catégories de résultats : les dépôts mesurables (**A**), les dépôts où on lit une erreur de mesure (**B**) et ceux qui ne sont pas mesurables (**C**). On constate dans le **Tableau C.36** que les données sont caractérisées par trois variables quantitatives : la valeur **x** de la cible utilisée, la durée du dépôt **t** et la température du substrat **T**. Cette ensemble de variables quantitatives est rattaché à une variable qualitative : la « mesurabilité » de la permittivité diélectrique (**ε**).

On veut donc savoir à présent pour les treize expériences du plan si elles vont conduire à un **ε** mesurable (**A**), à une erreur de mesure (**B**) ou à aucune mesure (**C**). Pour cela, on effectue une **A.F.D.** à partir des données du **Tableau C.36** auxquelles on ajoute les 13 expériences du plan d'expérimentations en tant qu'individus supplémentaires dans l'analyse.

Cette analyse nous permet donc de prévoir si l'on pourra mesurer la réponse ε pour toutes les expériences du plan d'expérimentations. Ainsi, nous obtenons une probabilité d'appartenance des individus aux différents groupes (**A**, **B** et **C**) qui conduit à un classement de ces individus (E1, E2, ..., E13) dans les différents groupes (voir **Tableau C.44**).

Tableau C.44 Classement, probabilités d'appartenance des individus supplémentaires

N°essai	Groupe	Prob. A	Prob. B	Prob. C
E1	A	0,753	0,189	0,058
E2	A	0,706	0,193	0,101
E3	A	0,969	0,020	0,011
E4	B	0,190	0,628	0,182
E5	B	0,205	0,651	0,144
E6	A	0,964	0,021	0,015
E7	A	0,832	0,103	0,066
E8	A	0,590	0,329	0,081
E9	A	0,611	0,327	0,062
E10	A	0,944	0,042	0,014
E11	A	0,808	0,104	0,088
E12	A	0,808	0,104	0,088
E13	A	0,732	0,192	0,076

Sur la **Figure C.41**, les triangles représentent les individus supplémentaires dont on recherche l'appartenance aux groupes **A**, **B** ou **C**. Ces groupes sont représentés par différentes couleurs : **A** (blanc), **B** (noir) et **C** (gris). Les triangles blancs représentent donc les individus supplémentaires classés dans le groupe **A** et les triangles noirs, ceux classés dans le groupe **B**. On constate donc conformément au **Tableau C.44**, qu'aucune des 13 expériences supplémentaires n'a été classée dans le groupe **C** (triangle gris) et que toutes les expériences sont classées dans le groupe **A** (mesurable) sauf deux (**E4** et **E5**) qui sont classées en **B** (erreur de mesure).

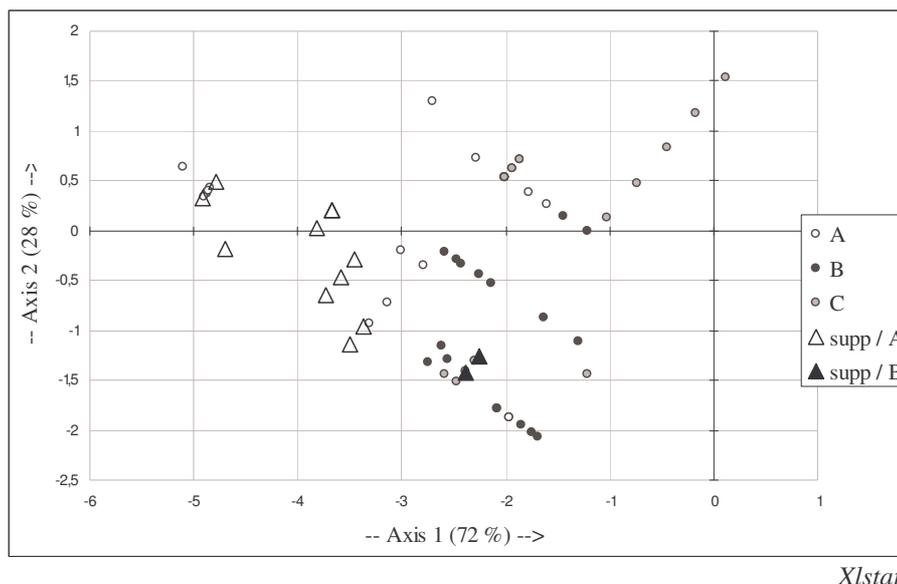


Figure C.41 Classement des individus dans les groupes de la variable mesure de ε

Ces résultats montrent que dans le plan d'expérimentations à venir, seules deux expériences risquent de conduire à une mauvaise mesure de la permittivité. **S. Liebus** ayant terminé ses travaux de thèse, le plan d'expérimentation prévu n'a pu être réalisé.

II.5. CONCLUSION

L'analyse des données nous a permis sans effectuer d'expériences supplémentaires, de trouver d'une part les corrélations existantes entre les différents paramètres d'une expérimentation (grâce à une **A.C.P**) et d'autre part de faire une prédiction judicieuse pour la suite des expériences (grâce à une **A.F.D**).

Quand à la méthodologie des plans d'expériences, dans ce cas la méthodologie des surfaces de réponses (plus précisément, un plan de **Doehlert**), elle nous permettra d'identifier de façon utile et structurée le réglage optimum d'une expérimentation pour obtenir une valeur de la permittivité maximale dans le cas où cette étude serait reprise.

Chapitre III : Application à la caractérisation hyperfréquence de couches minces

Ces travaux ont été menés en collaboration avec **A. Reynet** qui a soutenu sa thèse [7] en septembre 2004 au sein de l'équipe **Circuits et Dispositifs Microondes (CDM)** de l'Institut de **Recherche en Communications Optiques et Microondes (IRCOM)**. Le principal objectif de l'utilisation des plans d'expériences dans ce chapitre, est de caractériser la permittivité réelle d'une couche mince (**CM**).

III.1. FORMALISATION DU PROBLEME

III.1.1. OBJECTIF

III.1.1.a. **Circuit de caractérisation**

Un circuit spécifique a été développé pour caractériser des couches minces (CM) de façon non destructive. Les CM à étudier sont des dépôts de titanate de baryum et de strontium déposées par **S. Liebus** [6]. Les caractéristiques de la CM à mesurer sont : sa **permittivité réelle** et sa **tangente de pertes**. Le caractère non destructif de la méthode impose deux nouvelles variables à déterminer : l'épaisseur de la **lame d'air** entre le circuit et la couche mince et l'épaisseur de la **lame d'air** entre le circuit et le substrat de MgO sur lequel est déposée la couche mince. Cette épaisseur de lame d'air modélise les défauts de contact.

Une mesure en transmission du dispositif permet de remonter aux caractéristiques électromagnétiques (EM) du circuit, du substrat et de la couche mince. Pour cela, il est nécessaire de déterminer deux fréquences de résonance. Cette étude a été réalisée à l'aide d'un logiciel de simulation EM 2,5D.

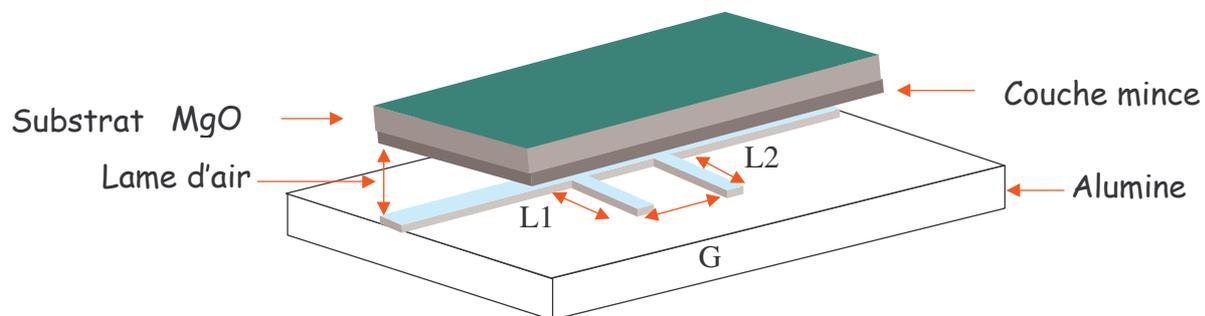
3 mesures successives sont effectuées :

- une première mesure du circuit à vide : **étape 1**,
- une deuxième mesure avec le substrat de MgO posé sur le circuit : **étape 2**,
- une dernière mesure avec le dépôt déposé sur le substrat : **étape 3**. Cette mesure s'effectue avec la couche mince posée sur le circuit.

La première mesure à vide permet de caractériser les paramètres EM du circuit (caractéristiques EM de l'alumine porteur du circuit et conductivité des lignes).

La deuxième mesure permet de caractériser le substrat : caractéristiques EM du MgO et épaisseur de la lame d'air qui modélise les défauts de contact entre le substrat de MgO et le circuit. La dernière mesure permet de caractériser la CM déposée sur le substrat préalablement caractérisé c'est-à-dire les caractéristiques EM de la CM et l'épaisseur de lame d'air entre la CM et le circuit.

La **Figure C.42** représente le capteur étudié pour la caractérisation des couches minces ferroélectriques. Le circuit est un substrat en alumine, de permittivité théorique 9,8 et d'épaisseur 380 μm sur lequel ont été déposées des lignes de largeur 350 μm et d'épaisseur 8 μm . Ce circuit est composée de 2 résonateurs en $\lambda/4$, l'un mesure 2,268 mm et l'autre 2,285 mm, l'espacement entre les deux résonateurs et de 5,597 mm (distance étudiée afin qu'il y ait un léger couplage entre les deux résonateurs).



Dimensions du circuit:

$L_1 = 2,268 \text{ mm}$

$L_2 = 2,285 \text{ mm}$

$G = 5,597 \text{ mm}$

➤ Hauteur Alumine : 380 μm

➤ Largeur de ligne : 350 μm

➤ Épaisseur de métallisation : 8 μm

Fréquences de résonance à vide:

✓ $f_1 = 12,80 \text{ GHz}$

✓ $f_2 = 13,34 \text{ GHz}$

Figure C.42 Capteurs double stubs

III.1.1.b. Circuit simulé

La simulation du circuit permet d'obtenir des valeurs de fréquences simulées (F_1 et F_2) avec les variables d'entrées correspondant aux paramètres EM du circuit, du substrat et de la couche mince. Nous aurons donc trois étapes de simulation pour obtenir les fréquences simulées du circuit à vide (**étape 1**), du substrat posé sur le circuit (**étape 2**) et de la CM posé sur le circuit (**étape 3**). Ces simulations (voir **Figure C.43**) sont réalisées à l'aide du logiciel **Momentum**.



Figure C.43 Schéma des simulations

Les valeurs des réponses F_1 et F_2 résultent ici d'une simulation numérique et devront être comparées aux valeurs $F_{1\text{exp}}$ et $F_{2\text{exp}}$ mesurées avec un analyseur de réseau vectoriel. Pour cela, l'expérimentateur fait varier les variables d'entrée avec une démarche proche de la méthode des essais et des erreurs jusqu'à faire coïncider les valeurs mesurées et les valeurs simulées. Il en déduit ainsi les bons réglages des facteurs d'entrées. Les facteurs d'entrée sont les propriétés EM du circuit, du substrat et du dépôt en sachant que le critère final recherché est la permittivité réelle ϵ_r de la couche mince (voir **chapitre II** de cette même partie) [6].

L'utilisation des plans d'expériences va avoir ici comme objectif la réduction du temps nécessaire à l'utilisateur pour optimiser le capteur étudié. Dans la méthodologie des plans d'expériences, on appelle les résultats de simulation, des « **tirs numériques** » pour marquer la différence avec le résultat d'un « mesurage » qui sera entachée d'une incertitude de mesure. Nous allons donc chercher à minimiser le nombre de tirs numériques et à obtenir des informations sur les réglages des variables d'entrées grâce à un plan d'expériences.

Le réglage des facteurs pour les différentes étapes (1, 2 et 3) va donc nécessiter la construction de trois plans d'optimisation successifs.

Objectif : « déterminer les paramètres EM des matériaux »

III.1.2. REPONSES

Les fréquences mesurées à l'aide d'un analyseur de réseau vectoriel [7] correspondent aux valeurs cibles des réponses que l'on souhaite obtenir avec la simulation du circuit. Les valeurs des deux fréquences de mesure du dispositif sont déterminées à partir de la **Figure C.44** .

La première fréquence ($F_{1\text{exp}}$) en (GHz) est obtenue avec le premier pic de résonance et la deuxième ($F_{2\text{exp}}$) avec la mesure du deuxième pic.

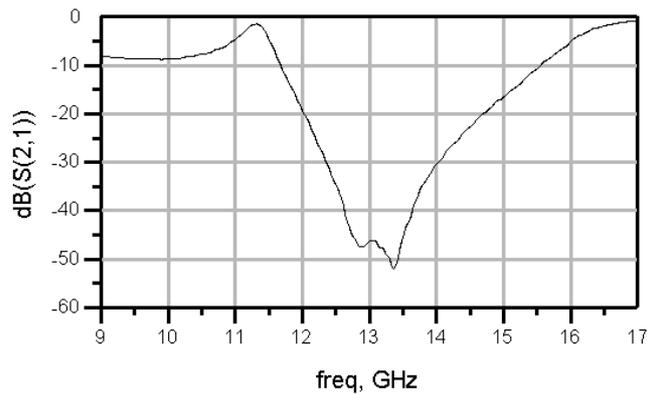


Figure C.44 Mesure en transmission du dispositif

A chacune des étapes de caractérisation, une mesure en transmission du dispositif est effectuée afin d'obtenir les fréquences de résonance de chaque dispositif. Nous allons exposer ci-dessous les valeurs des réponses pour chacune de ces étapes.

III.1.3. FACTEURS

Les facteurs de l'étude sont reportés dans le **Tableau C.45** ci-dessous.

Tableau C.45 Facteurs des étapes

N° étape	Facteur 1	Facteur 2	Facteur 3
Etape 1	Permittivité réelle du substrat du circuit : ϵ	Tangente de pertes du substrat du circuit : $\tan \delta$	Conductivité des lignes : σ
Etape 2	Permittivité réelle du substrat porteur : ϵ_{sub}	Epaisseur de lame d'air entre circuit et substrat : L_{airsub}	Tangente de pertes du substrat porteur : $\tan \delta_{\text{sub}}$
Etape 3	Permittivité réelle de la CM : ϵ_{CM}	Epaisseur de lame d'air entre circuit et CM : L_{airCM}	Tangente de pertes de la CM : $\tan \delta_{\text{CM}}$

III.2. ETAPE 1 : PARAMETRES EM DU CIRCUIT A VIDE

III.2.1. DESCRIPTION DU PLAN D'EXPERIENCES DE L'ETUDE 1

Le premier plan d'expériences consiste à obtenir les réponses simulées (F_1 et F_2) proches des réponses mesurées ($F_{1\text{exp}}$ et $F_{2\text{exp}}$) afin de déterminer les paramètres EM du circuit à vide.

Par soucis de présentation et pour ne pas surcharger ce chapitre, nous allons résumer les objectifs dans le **Tableau C.46**, les niveaux des facteurs dans le **Tableau C.47**, les réponses dans le **Tableau C.48** ainsi que le plan d'expérimentations et les résultats d'essais dans le **Tableau C.49**.

Tableau C.46 Objectif de l'étude de l'étape 1

Objectif de l'étude	Surface de Réponses / Matrice de Doelhart
Nombre de variables	3
Nombre d'expériences	13
Nombre de coefficients	10
Nombre de réponses	2

Tableau C.47 Facteurs et niveaux de l'étape 1

	Facteur	Unité	Umin	Umax
U1	tan perte	m U	0	1
U2	permittivité		8	12
U3	conductivité	MegaS.	20	80

Tableau C.48 Réponse de l'étape 1

	Réponses simulées	Unité
Y1	F1	GHz
Y2	F2	GHz

Tableau C.49 Plan d'expérimentations et résultats d'essais de l'étape 1

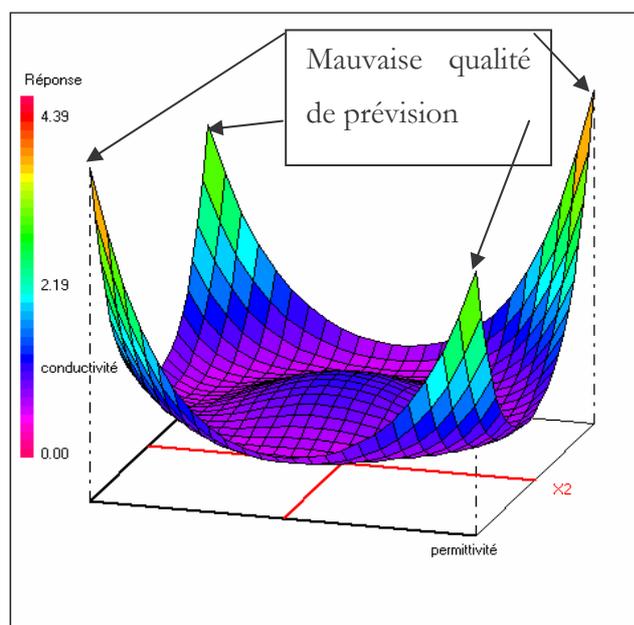
N°Exp	Facteurs			Réponses	
	tan perte	permittivité	conductivité	F1	F2
1	1	10	50	12,555	12,875
2	0	10	50	12,55	12,876
3	0,75	12	50	11,513	11,819
4	0,25	8	50	13,927	14,312
5	0,75	8	50	13,911	14,329
6	0,25	12	50	11,51	11,817
7	0,75	10,667	80	12,177	12,5
8	0,25	9,333	20	12,991	13,286
9	0,75	9,333	20	13	13,284
10	0,5	11,333	20	11,835	12,117
11	0,25	10,667	80	12,173	12,498
12	0,5	8,667	80	13,416	13,792
13	0,5	10	50	12,551	12,875

III.2.2. ANALYSES ET RESULTATS

III.2.2.a. Qualité prévisionnelle du modèle

On rappelle que l'utilisation d'un modèle mathématique doit conduire à la valeur de la réponse étudiée en n'importe quel point du domaine expérimental (dans ce cas F_1 et F_2). Cette valeur que nous appelons **prévision**, n'est pas égale à la véritable valeur (voir **Partie A**) mais à une estimation de cette véritable valeur ($y_i = \hat{y}_i + e_i$). La qualité de cette prévision représente la proximité entre l'estimation et la véritable valeur. Pour tester cette qualité de prévision, on calcule la variance de la réponse calculée ($\text{Var}(\hat{y}_i)$) pour un ensemble de points répartis dans le domaine expérimental.

Nous avons choisi ici des matrices de **Doelhart**. Rappelons, que ces matrices s'utilisent dans un domaine expérimental sphérique avec lequel nous explorons une zone du plan autour de son centre défini comme point de référence. Il faudra donc faire attention aux conclusions sur les résultats obtenus à l'aide de la réponse calculée se situant aux extrémités du domaine expérimental.



La **Figure C.45** représente la courbe d'isovariance par rotation en 3D (voir **Partie A**) valable quelle que soit la réponse issue du plan de réseaux de **Doelhart** pour les facteurs conductivité et permittivité.

D'après la **Figure C.45**, on voit que la fonction de variance n'a pas la même valeur au sommet du domaine qu'au centre. Il faudra donc faire attention à la qualité de prévision des valeurs extrêmes du domaine expérimental.

Figure C.45 Courbe d'isovariance par rotation (conductivité, permittivité)

III.2.2.b. Validité du modèle

L'analyse statistique aboutissant à la validité du modèle est décrite dans un **tableau d'analyse de régression**. Nous n'avons pas de répétitions étant donné que nous faisons des simulations : en effet la simulation d'une combinaison de facteurs répétées donne la même réponse. Ainsi, nous n'avons pas pu faire le **test de validation 2 (Partie A)**. Les résultats des **tests de validation 1** pour la réponse F_1 (**Tableau C.50**) et F_2 (**Tableau C.51**) sont très **significatifs**. En effet, la probabilité de rejeter l'hypothèse que le modèle ne permet pas de décrire la variation des résultats d'essais (voir **partie A**), est inférieure à 0,01.

Tableau C.50 Tableau d'analyse de régression pour la réponse F_1

Source de variation	Somme des carrés	ddl	Carré moyen	Rapport	Probabilité
Régression	7,774	9	0,8638	25758,1394	< 0,01 ***
Résidus	0,0001	3	0		
Total	7,7741	12			

Tableau C.51 Tableau d'analyse de régression pour la réponse F_2

Source de variation	Somme des carrés	ddl	Carré moyen	Rapport	Probabilité
Régression	8,3698	9	0,93	5478,5963	< 0,01 ***
Résidus	0,0005	3	0,0002		
Total	8,3703	12			

III.2.2.c. Analyses graphiques

Nous allons analyser dans un premier temps les surfaces de réponses de la réponse F_1 (Figure C.46) et de F_2 (Figure C.47).

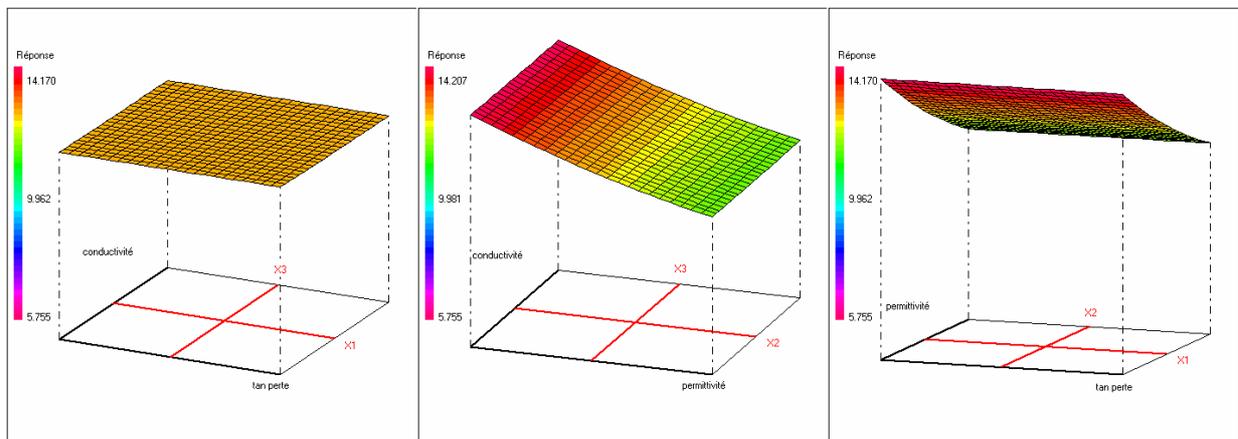


Figure C.46 Surface de réponses de F_1

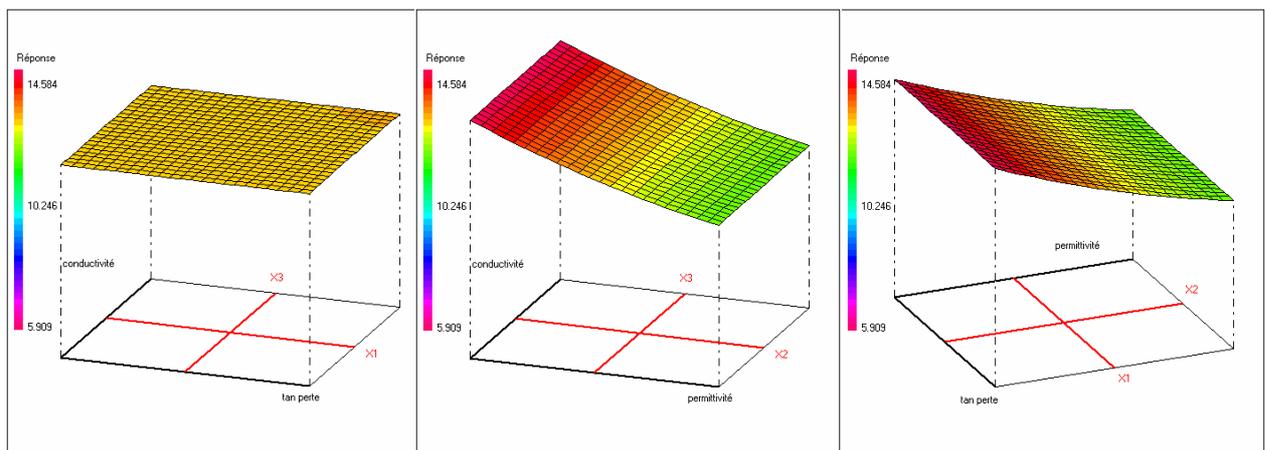


Figure C.47 Surface de réponses de F_2

En observant ces figures, nous pouvons nous interroger sur le degré du modèle choisi (degré 2), les surfaces de réponses semblent être planaires. Pouvons nous décrire le phénomène avec un modèle de degré 1 ? Pour le savoir, nous allons regarder la signification des coefficients du modèle de degré 2 choisi (Tableau C.52).

Tableau C.52 Tableau d'analyse des coefficients

Nom	Coefficient	Signif, %
b0	12,551	< 0,01 ***
b1	0,0012	69,3
b2	-1,3891	< 0,01 ***
b3	-0,0122	2,25 *
b11	0,0015	83,9
b22	0,2185	< 0,01 ***
b33	0,0165	9,6
b12	0,011	19,9
b13	-0,0069	42,4
b23	0,0198	7,6

Les résultats montrent qu'un modèle de degré inférieur ne peut pas être choisi. En effet, le coefficient b_{22} associé au modèle de degré 2 détient un pourcentage d'être nul très faible (<0,01%). Donc, ce coefficient est significatif.

Le modèle de degré 2 a donc été retenu pour les études ultérieures.

Dans cette étude, nous cherchons à trouver les valeurs des facteurs d'entrées qui correspondent aux valeurs cibles F_{1exp} et F_{2exp} . Ceci revient à recouper les six graphiques (**Figure C.46** et **Figure C.47**) pour trouver les valeurs de la conductivité, de la permittivité et de la tangente de pertes. Cette étude revient à trouver un optimum multi-critères et donc à faire un compromis entre les deux réponses F_1 et F_2 .

III.2.2.d. Optimum multicritère : fonction de désirabilité ⁽¹⁾

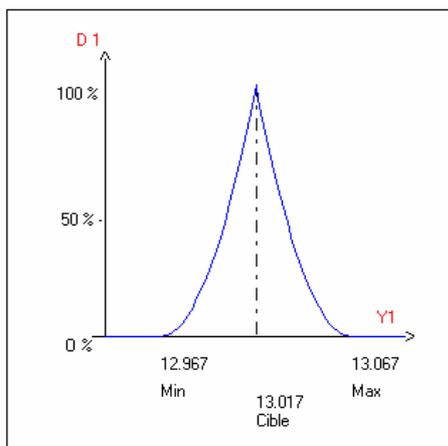
Il s'agit ici de trouver une méthode de recherche d'un compromis entre plusieurs réponses. L'utilisation du modèle pour la recherche d'une solution nécessite une transformation des réponses afin de les rendre comparables. On utilise ici les fonctions de désirabilité qui facilitent, grâce à des représentations graphiques simples, l'interprétation de ces transformations mathématiques.

La recherche d'un optimum multi-critères avec la fonction désirabilité se fait en trois phases :

- la **première phase** consiste à transformer chacune des réponses en une fonction de désirabilité individuelle dont la nature dépend des objectifs de l'étude : la recherche d'un minimum, la recherche d'un maximum ou encore la recherche d'une valeur particulière (valeur cible). Cette transformation donne lieu à une représentation graphique,
- la **deuxième phase** consiste à définir une fonction de désirabilité globale à partir de la moyenne géométrique des fonctions de désirabilité individuelle. On peut pondérer les fonctions de désirabilité individuelle dans le calcul de la désirabilité globale pour tenir compte de l'importance relative des réponses,

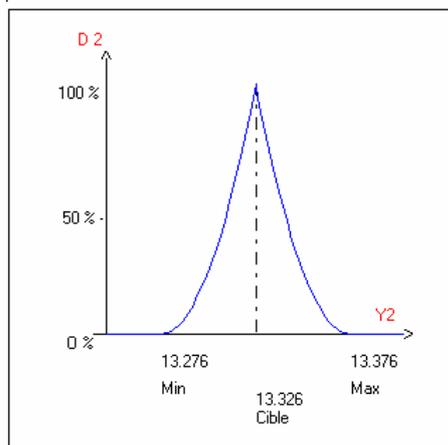
- la **dernière phase** consiste à rechercher le niveau des facteurs qui permet d'atteindre le maximum de la fonction de désirabilité globale. Toute valeur strictement positive de la fonction de désirabilité globale traduit un réglage particulier des facteurs permettant d'atteindre, à partir des modèles empiriques, des valeurs des réponses à l'intérieur des intervalles de tolérance. Cette recherche se fait de manière itérative dans les logiciels.
- **Phase 1 : détermination des fonctions de désirabilité individuelle**

La transformation d'une valeur particulière d'une réponse modélisée en un indice de satisfaction est illustrée de manière graphique ci-dessous (**Figure C.48** et **Figure C.49**)



La **Figure C.48** représente la fonction de désirabilité (D_1) de la fréquence simulée F_1 . Ici, nous cherchons une valeur cible qui permet d'avoir la fréquence simulée égale à la fréquence mesurée F_{1exp} (13,017 GHz). On a pris un intervalle de tolérance avec un pas de 0,05 GHz ce qui donne une fréquence désirée minimale de 12,967 GHz et une fréquence désirée maximale de 13,067 GHz.

Figure C.48 Fonction désirabilité de la réponse F_1



La **Figure C.49** représente la fonction de désirabilité (D_2) de la fréquence simulée F_2 . Nous recherchons à nouveau une valeur cible F_{2exp} (13,326 GHz). Les fréquences désirées minimale et maximale sont respectivement égales à 12,967 GHz et 13,067 GHz.

Figure C.49 Fonction désirabilité de la réponse F_2

Les réponses modélisées ont donc été transformées en fonctions de désirabilité variant de 0 à 100%. La **désirabilité** représente le degré de satisfaction des expérimentateurs en fonction de la valeur obtenue de la réponse modélisée après lui avoir donné une valeur cible à atteindre. On attribue un indice de satisfaction égal à 100% lorsque l'objectif est atteint. Cet indice est égal à zéro si la valeur de la réponse modélisée se situe à l'extérieur de l'intervalle de tolérance qu'on lui a associé. Toute valeur située à l'intérieur de l'intervalle de tolérance fait l'objet d'une transformation :

$$\left\{ \begin{array}{ll} D = \left(\frac{\eta_i - \eta_{\min}}{\eta_{\text{cible}} - \eta_{\min}} \right)^s & \text{si } \eta_{\min} \leq \eta_i < \eta_{\text{cible}} \\ D = 1 & \text{si } \eta_i = \eta_{\text{cible}} \\ D = \left(\frac{\eta_i - \eta_{\max}}{\eta_{\text{cible}} - \eta_{\max}} \right)^t & \text{si } \eta_{\max} \geq \eta_i > \eta_{\text{cible}} \end{array} \right.$$

avec : η_i : réponse modélisée ou réponse calculée,

η_{\min} : réponse minimale modélisée désirée de l'intervalle de tolérance,

η_{\max} : réponse maximale modélisée désirée de l'intervalle de tolérance,

η_{cible} : réponse cible désirée,

s : coefficient gauche de l'évolution non linéaire de la fonction de désirabilité,

t : coefficient droit de l'évolution non linéaire de la fonction de désirabilité.

La **Figure C.50** illustre graphiquement la signification des coefficients s et t. Dans l'étude, nous avons choisi des coefficients s et t égaux à 2 de manière à se rapprocher le plus possible de la valeur cible.

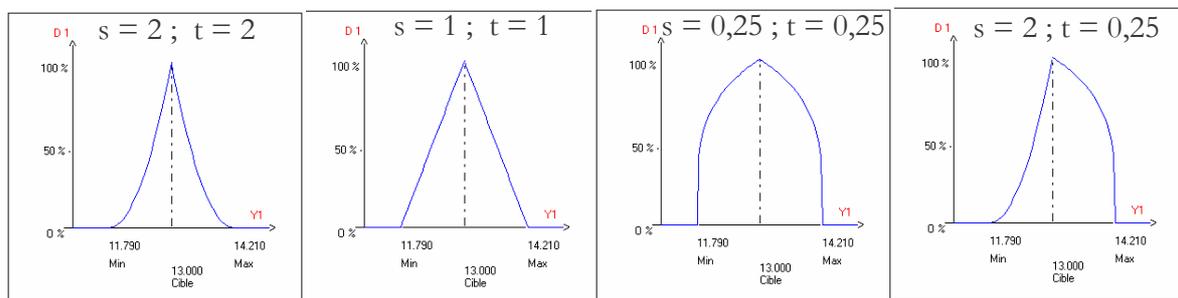


Figure C.50 Fonction de désirabilité de la réponse F_1 à différent coefficient s et t

▪ Phase 2 : fonction de désirabilité globale

Après avoir transformé toutes les réponses en fonctions de désirabilité individuelle, il est nécessaire de définir un critère objectif que l'on va ensuite chercher à optimiser.

La valeur de la fonction de désirabilité globale est définie à partir de la moyenne géométrique des valeurs des fonctions de désirabilité individuelle. Il est également possible d'accorder un poids aux différentes réponses. Dans le cas présent, on obtient :

$$D = (D_1^{\text{Poids1}} \times D_2^{\text{Poids2}})^{\frac{1}{\text{poids1} + \text{poids2}}}$$

avec D_1 : fonction de désirabilité individuelle de la réponse F_1 ,

D_2 : fonction de désirabilité individuelle de la réponse F_2 ,

Poids1 : poids affecté à la réponse F_1 (de 1 à 10),

Poids2 : poids affecté à la réponse F_2 (de 1 à 10).

Dans notre cas, on obtient : $D = \sqrt{D_1 \times D_2}$ car nous avons attribué le même poids au deux réponses (poids1 = poids2 =1).

Remarque : dès que le niveau d'une des réponses n'appartient pas à l'intervalle de tolérance qui lui est associé, la valeur de la fonction de désirabilité globale **D** est nulle.

▪ **Phase 3 : recherche de l'optimum multi-critères**

La recherche d'un optimum multi-critères consiste à trouver le niveau des facteurs qui maximise la valeur de la fonction de désirabilité globale. Pour cela, on procède, grâce à des logiciels, à des itérations des combinaisons. Le **Tableau C.53** présente les résultats obtenus avec le logiciel **NemrodW**.

Tableau C.53 Résultat de la recherche de l'optimum

Facteur	Valeur	
tan perte	0,344	
permittivité	9,291	
conductivité	26	
Nom de la réponse	Valeur	di %
F1	13,017	98,21
F2	13,325	94,61
DESIRABILITE D		96,39

Nous obtenons un fort pourcentage de désirabilité (noté **di**), pour les deux fréquences F_1 (98,21%) et F_2 (94,61%) ce qui conduit à un pourcentage élevé de la désirabilité globale (96,39%).

Ainsi, nous avons pu remonter aux paramètres EM du circuit (tangente de pertes, permittivité et conductivité).

Le résultat concernant la permittivité est en accord avec des études précédemment menées par **A. Reynet** sur ce circuit [7]. A l'inverse, le résultat obtenu pour la conductivité semble diverger de l'étude précédemment citée.

Pourquoi le plan donne t il une valeur divergente ?

Comme nous l'avons dit précédemment (**III.2.2.a.**), la qualité prévisionnelle du modèle sera mauvaise en limite du domaine expérimental. Or, dans le cas de la conductivité, le modèle donne une valeur égale à 26 qui est proche de la valeur U_{\min} du domaine expérimental ($U_{\min} = 20$).

III.3. ETAPE 2 : PARAMETRES EM DU SUBSTRAT

III.3.1. DESCRIPTION DU PLAN D'EXPERIENCES DE L'ETUDE 2

Le deuxième plan d'expériences consiste à obtenir les réponses simulées (F_1 et F_2) proches des réponses mesurées ($F_{1\text{exp}}$ et $F_{2\text{exp}}$) afin de déterminer les paramètres EM du substrat.

Les objectifs (**Tableau C.54**), les niveaux des facteurs (**Tableau C.55**), les réponses (**Tableau C.56**) ainsi que le plan d'expérimentations et les résultats des essais (**Tableau C.57**) sont résumés ci-dessous.

Tableau C.54 Objectif de l'étude de l'étape 2

Objectif de l'étude	Surface de Réponses / Matrice de Doelhart
Nombre de variables	3
Nombre d'expériences	13
Nombre de coefficients	10
Nombre de réponses	2

Tableau C.55 Facteurs et niveaux de l'étape 2

	Facteur	Unité	Umin	Umax
U1	lame d'air	μm	16	32
U2	permittivité		8,2	11
U3	tan perte	mU	0	10

Tableau C.56 Réponse de l'étape 2

	Réponses simulées	Unité
Y1	F1	GHz
Y2	F2	GHz

Tableau C.57 Plan d'expérimentations et résultats d'essais de l'étape 2

N°Exp	Facteurs			Réponses	
	lame d'air	permittivité	tan perte	F1	F2
1	32	9,6	5	12,034	12,416
2	16	9,6	5	11,645	11,972
3	28	11	5	11,88	12,242
4	20	8,2	5	11,89	12,263
5	28	8,2	5	12,062	12,396
6	20	11	5	11,668	12,018
7	28	10,0667	10	11,935	12,3
8	20	9,1333	0	11,802	12,146
9	28	9,1333	0	11,983	12,36
10	24	10,5334	0	11,809	12,17
11	20	10,0667	10	11,742	12,08
12	24	8,6666	10	11,942	12,3
13	24	9,6	5	11,874	12,225

III.3.2. ANALYSES ET RESULTATS

Nous avons procédé à la même recherche de désirabilité que dans l'étape 1. Dans cette analyse, seuls les critères cibles des fréquences changent. A présent, on désire obtenir :

$$F_{1\text{exp}} = 11,895 \text{ GHz} ; F_{2\text{exp}} = 12,089 \text{ GHz}.$$

Les fréquences de résonances mesurées sont entrées dans le logiciel **NemrodW** afin de permettre à celui-ci de converger vers les paramètres du substrat de MgO, on obtient :

épaisseur de lame d'air : 22 μm ,

permittivité du MgO : 9,58,

tangente de pertes : $5 \cdot 10^{-3}$.

La valeur de la permittivité obtenue avec le plan d'expériences est très proche de la valeur tabulée dans la littérature ($9,6 < \epsilon_r < 10$ [7]) ce qui montre la validité de ce plan, ainsi que sa qualité prédictive.

III.4. ETAPE 3 : PARAMETRES EM DE LA CM

III.4.1. DESCRIPTION DU PLAN D'EXPERIENCES DE L'ETUDE 3

Le dernier plan d'expériences consiste maintenant à obtenir les réponses simulées (F_1 et F_2) proches des réponses mesurées ($F_{1\text{exp}}$ et $F_{2\text{exp}}$) afin de déterminer les paramètres EM de la CM.

Les objectifs (**Tableau C.58**), les niveaux des facteurs (**Tableau C.59**), les réponses (**Tableau C.60**) ainsi que le plan d'expérimentations et les résultats des essais (**Tableau C.61**) sont résumés ci dessous.

Tableau C.58 Objectif de l'étude de l'étape 3

Objectif de l'étude	Surface de Réponses / Matrice de Doelhert
Nombre de variables	3
Nombre d'expériences	13
Nombre de coefficients	10
Nombre de réponses	2

Tableau C.59 Facteurs et niveaux de l'étape 3

	Facteur	Unité	Umin	Umax
U1	lame d'air	μm	16	32
U2	permittivité		200	600
U3	tan perte	mU	0	10

Tableau C.60 Réponse de l'étape 3

	Réponses simulées	Unité
Y1	F1	GHz
Y2	F2	GHz

Tableau C.61 Plan d'expérimentations et résultats d'essais de l'étape 3

N°Exp	lame d'air	Facteurs		Réponses	
		permittivité	tan perte	F1	F2
1	32	400	5	11,973	12,346
2	16	400	5	11,509	11,881
3	28	600	5	11,859	12,212
4	20	200	5	11,721	12,069
5	28	200	5	11,927	12,278
6	20	600	5	11,627	11,976
7	28	467	10	11,884	11,231
8	20	333	0	11,686	12,042
9	28	333	0	11,899	12,26
10	24	533	0	11,757	12,128
11	20	467	10	11,652	12,008
12	24	267	10	11,816	12,171
13	24	400	5	11,79	12,144

III.4.2. ANALYSES ET RESULTATS

Nous avons procédé à la même recherche de désirabilité que dans les étapes précédentes. Dans cette analyse, les critères cibles des fréquences à obtenir sont :

$$F_{1\text{exp}} = 12,043 \text{ GHz} ; F_{2\text{exp}} = 12,209 \text{ GHz}$$

L'étude de la désirabilité permet alors, de remonter aux caractéristiques suivantes (résultat du logiciel **NemrodW**) :

$$\varepsilon_r = 350,$$

$$\tan\delta = 5.10^{-3},$$

$$\text{lame d'air} = 32 \mu\text{m}.$$

Des études exposées dans la thèse de **S. Liebus [6]** montrent que pour un même matériau, des résultats équivalents sur la permittivité réelle sont obtenus.

III.5. CONCLUSION

Dans cette étude, trois plans d'optimisations menés avec des réseaux uniformes de **Doelhart**, nous ont permis de trouver avec une bonne précision les valeurs des permittivités du circuit, du substrat et de la couche mince. Par ailleurs, les valeurs de certains facteurs trouvées après une étude de désirabilité au bord du domaine expérimental, nous laissent penser que l'utilisation de

matrices dont le domaine expérimental serait cubique, pourrait peut être mieux correspondre au problème.

Dans cette étude, les plans d'expériences ont permis à l'expérimentateur d'effectuer beaucoup moins de tirs numériques et ainsi de gagner un temps utilisateur très important. En effet, la détermination des caractéristiques EM d'une couche mince nécessitait un temps utilisateur d'un mois et demi, avec interaction continue de l'utilisateur avec le logiciel. L'utilisation des plans d'expériences a fortement réduit les temps de calcul ainsi que les temps d'interaction de l'utilisateur avec le logiciel. A savoir, un temps de calcul de 3 semaines pour un temps d'interaction avec le logiciel d'une quinzaine d'heures. De plus, cette étude a permis de valider le circuit utilisé pour caractériser des couches minces à forte permittivité.

Grâce à cet exemple, nous avons ajouté à la présentation de l'utilisation des plans d'expériences pour l'étude des surfaces de réponse, des éléments complémentaires en ce qui concerne une méthode d'optimisation multi-critères.

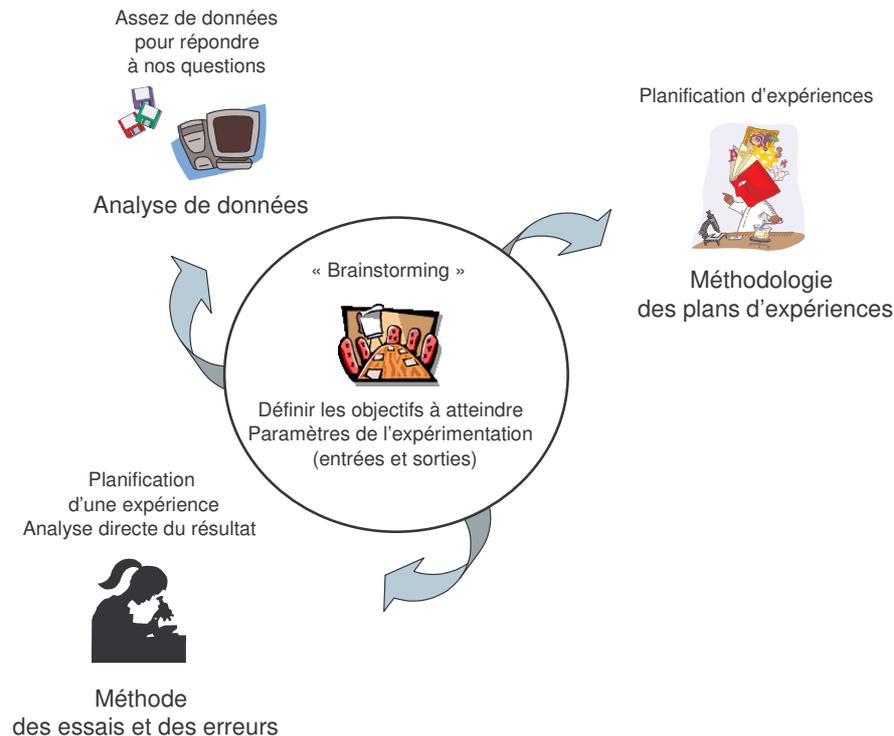
Bibliographie Partie C

- [1] H. Hidalgo, Dépôt chimique en phase vapeur de couches minces d'alumine dans une post-décharge micro-onde, *Thèse de doctorat en Chimie, Chimie Physique, Spécialité Procédés et Céramiques*, Université de Limoges, septembre 2003.
- [2] T. Haure, Couches multifonctionnelles par procédé multitechnique, *Thèse de doctorat en Chimie, Chimie Physique, Spécialité Procédés et Céramiques*, Université de Limoges, novembre 2003.
- [3] J.P. Burmann, R.L. Plackett, *The design of optimum multifactorial experiments*, *Biometrika*, 1946, Vol 33, p 305 – 325.
- [4] J. Alexis, P. Alexis, *Pratique industrielle des plans d'expériences – La qualité à moindre coût : l'approche Taguchi*, Afnor, Ed. Paris, 1999.
- [5] D. Benoist, Y. Tourbier, S. Germain-Tourbier, *Plans d'expériences : construction et analyse*, Technique & documentation- Ed. Lavoisier, 1994.
- [6] S. Liebus, Couches minces ferroélectriques appliquées aux dispositifs microondes accordables en fréquence, *Thèse de doctorat en Electronique des Hautes Fréquences et Optoélectronique*, Université de Limoges, novembre 2003.
- [7] A. Reynet, Contribution à la caractérisation micro onde de matériaux non magnétiques en couche mince par résonateur micro ruban et coaxiaux, *Thèse de doctorat en Electronique des Hautes Fréquences et Optoélectronique*, Université de Limoges, septembre 2004.

Conclusion générale

L'objectif de ce travail de thèse était de présenter, de comprendre et d'utiliser un ensemble de méthodes et de modes de raisonnement destinés à tout expérimentateur désirant faire de la planification expérimentale et/ou de l'analyse de données dans le but d'optimiser l'efficacité de sa recherche expérimentale. Il existe des méthodes pertinentes et judicieuses adaptées à chaque problématique.

Les différentes méthodes utilisées et leurs interactions sont résumées sur la figure ci-dessous.



La méthode des essais et des erreurs, qui consiste à effectuer des expériences de manière séquentielle en faisant varier les facteurs d'entrée du système étudié l'un après l'autre sans planification préalable, présente un certain nombre de limites. Même si cette méthode donne des résultats, elle est très coûteuse en temps car elle nécessite inévitablement la réalisation d'un grand nombre d'expériences.

En regroupant les acteurs d'une étude au cours d'un brainstorming, la réflexion collective doit permettre de trouver les causes possibles d'un problème et les solutions de celui-ci en terme de méthodes ou de combinaisons de méthodes à appliquer.

La méthodologie des plans d'expériences va aider l'expérimentateur à exprimer au mieux son problème en lui proposant des stratégies expérimentales optimales (enchaînements de plans d'expériences dans le temps) en fonction des objectifs qu'il s'est fixé et des moyens dont il dispose. Les objectifs peuvent être d'explorer un domaine expérimental inconnu (recherche

exploratoire dans le domaine de variation d'un ensemble de facteurs), d'isoler les facteurs influents (plan de criblage), d'élaborer des modèles descriptifs ou prévisionnels des phénomènes étudiés (étude quantitative des facteurs, étude quantitative des réponses), d'effectuer des optimisations (plans à surfaces de réponses).

Les méthodes d'analyses de données ne peuvent être mise en œuvre que si l'expérimentateur dispose d'un nombre suffisant de données expérimentales. L'objet de toutes ces méthodes est la description des relations entre plusieurs ensembles de données. Nous avons montré qu'à partir de représentations graphiques (cercles des corrélations), l'Analyse en Composantes Principales permet de visualiser les corrélations existantes entre les paramètres de l'expérimentation. Cette méthode ne s'applique qu'à des ensembles de données quantitatives. Si on s'intéresse aux relations entre une variable qualitative et un ensemble de variables quantitatives, on utilisera l'Analyse Factorielle Discriminante. Si on souhaite regrouper un grand nombre d'expériences en catégories jugées homogènes au regard de tel ou tel critère, on utilisera la Classification Hiérarchique Ascendante.

Les exemples traités dans ce manuscrit ont permis de montrer l'intérêt et la complémentarité de toutes ces méthodes.

Le premier cas étudié (optimisation des couches minces d'alumine réalisées par PECVD) en est un bon exemple. En effet, tout d'abord, un plan de criblage a été réalisé afin d'identifier les paramètres d'entrée importants ainsi que leur poids sur le contrôle du processus et les propriétés des films. Ensuite, un plan d'expériences pour l'étude des surfaces de réponses a permis d'obtenir un réglage plus affiné de ces paramètres correspondant à une optimisation des propriétés des films. Ces deux plans d'optimisation ont permis de déterminer les conditions optimales pour obtenir des couches de bonnes qualités. Il nous a ensuite semblé intéressant d'analyser l'ensemble des données générées par ces deux plans. Plusieurs Analyses en Composantes Principales ont été réalisées afin de déterminer tout d'abord les corrélations entre les propriétés des films puis, entre ces propriétés et les données du plasma et enfin, entre ces propriétés et la morphologie des splats.

Dans le deuxième exemple (optimisation des couches minces de titanate de baryum et de strontium obtenues par ablation laser), compte tenu des données déjà existantes, nous avons commencé par réaliser une Analyse en Composantes Principales puis une Classification Ascendante Hiérarchique. L'A.C.P. a permis de dégager des conclusions sur les influences des facteurs d'entrée et la C.A.H. de regrouper les expériences en familles (groupes homogènes). De plus, la mesure expérimentale de la réponse n'étant pas toujours faisable, une Analyse Factorielle Discriminante a montré qu'il était possible de faire une prédiction judicieuse pour la suite des

expériences. Cependant, ces analyses ne permettant pas d'obtenir les réglages exacts des paramètres pour optimiser la réponse, nous avons construit un plan d'optimisation qui n'a pas encore été réalisé.

Il nous a également semblé intéressant d'appliquer ces méthodes à l'optimisation de simulations numériques. Ce troisième exemple nous a conduit à utiliser une méthode d'optimisation multicritères. Les résultats obtenus ont permis de valider un circuit spécifique développé pour caractériser des couches minces de façon non destructive et de réduire considérablement les temps de calcul ainsi que le temps d'interaction de l'utilisateur avec le logiciel.

En conclusion dans ce manuscrit, nous avons développé et appliqué les méthodes qui nous semblaient les plus appropriées aux différents exemples étudiés. Les outils d'analyse et de synthèse et les plans d'expériences pouvant être utilisés pour traiter un problème sont nombreux et il est nécessaire de les adapter à chaque problématique. Les perspectives sur ces travaux seront d'approfondir la connaissance de ces méthodes mais également de découvrir et d'appliquer de nouveaux outils comme par exemple la méthode du simplexe ou encore la construction de plans optimaux (matrice à trous, construction de matrices particulières).

Annexes

Annexe 1 : Etude en cours

Optimisation des propriétés du niobate de lithium massif sous irradiation par faisceau d'électrons pilotés

Un des axes de recherche de l'équipe **C2M**, porte sur la réalisation de composants et dispositifs intégrés microoptiques. Dans ce cadre, nous avons travaillé sur l'optimisation des propriétés de structures à base de cristaux photoniques, réalisées sur niobate de lithium (LiNbO_3) et obtenues par une technique d'irradiation par faisceau d'électrons. Le faisceau d'électrons est délivré par un Microscope Electronique à Balayage (M.E.B.) qui permet la réalisation et l'observation des structures bidimensionnelles, de dimensions nanométriques (quelques centaines de nm). Le domaine d'applications visées est celui des Télécommunications Optiques. On cherche à optimiser les paramètres expérimentaux afin d'obtenir des cristaux photoniques ayant une bande interdite totale pour $\lambda = 1,55 \mu\text{m}$. Tous les résultats expérimentaux présentés dans ce chapitre font l'objet des travaux de la première année de thèse de **F. Quentin** [1].

I.1. FORMALISATION DU PROBLEME

I.1.1. DISPOSITIF EXPERIMENTAL

Une brève description du microscope électronique, des principaux paramètres d'irradiation utilisés pour la réalisation des structures et des caractéristiques recherchées va être présentée.

Le principe général du microscope électronique à balayage est représenté sur la **Figure I**. Il est basé sur le balayage ligne par ligne d'un faisceau d'électrons focalisés sur la surface d'un échantillon. En effectuant le même type de balayage en synchronisme sur l'écran d'observation, on visualise l'image de la surface de l'échantillon. Dans le cadre de cette étude, le pilotage du faisceau d'électrons primaire est pris en charge par un logiciel de lithographie (ELPHY) qui permet la conception des structures bidimensionnelles ainsi que leur réalisation. Il s'agit ici d'utiliser le M.E.B. non plus en balayant l'échantillon mais en déplaçant le spot électronique point par point, en mode "pulsé".

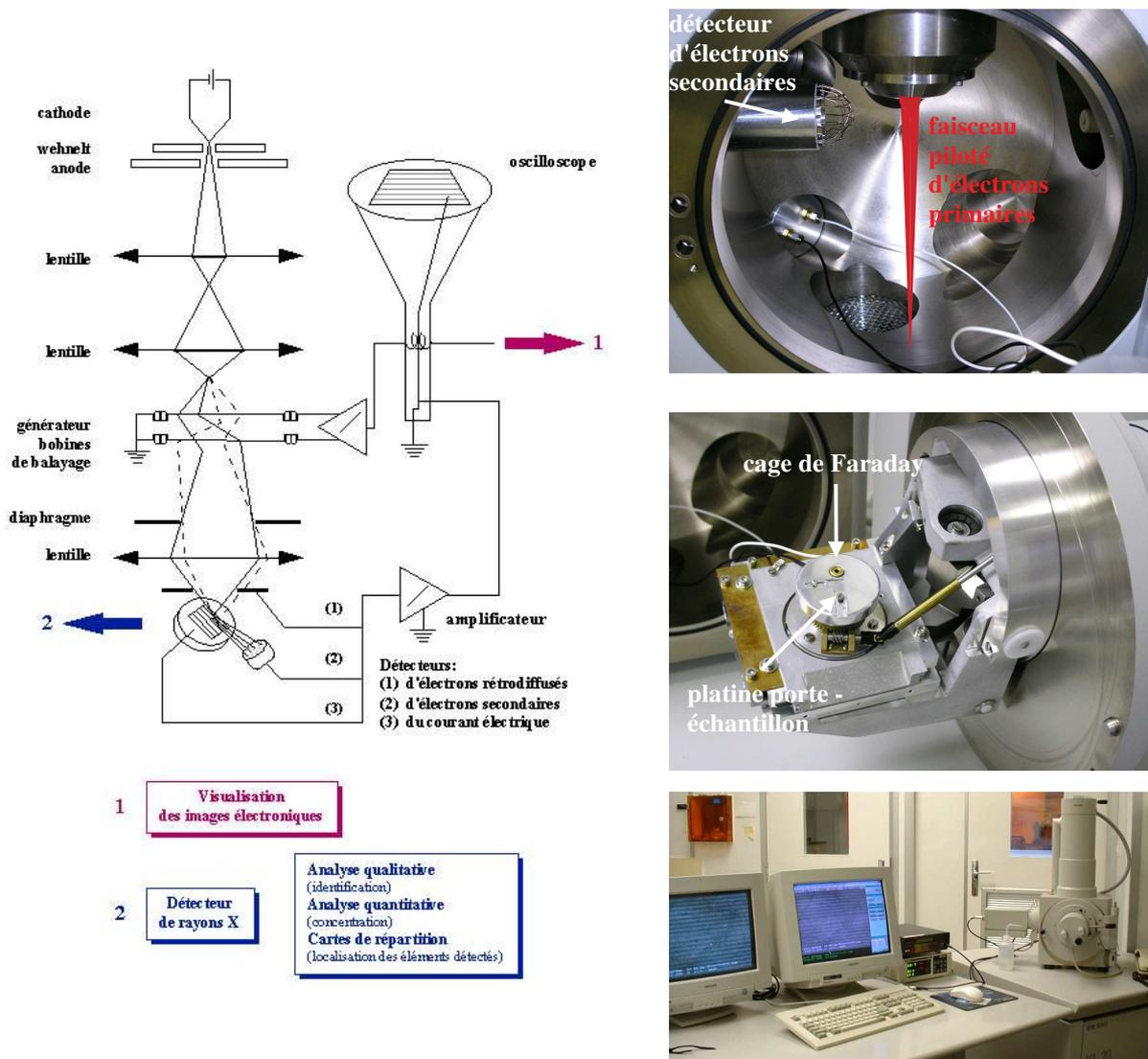


Figure I Visualisation et principe général du microscope électronique

Le procédé technologique utilisé comporte plusieurs étapes schématisées sur la **Figure II**. On commence par déposer, par évaporation, une couche d'or/palladium sur les deux faces du niobate de lithium. Sur la face z^+ , le dépôt métallique servira d'électrode et sur la face z^- , il évitera l'accumulation de charges pendant l'irradiation (**figure II (a)**). La seconde étape est la formation de trous d'air par bombardement électronique de la face z^- (**figure II (b)**) et la dernière étape, la révélation des zones irradiées par attaque acide $\text{HF}-\text{HNO}_3$ (**figure II (c)**). Des trous d'air apparaissent car la vitesse d'attaque est plus importante sur les zones irradiées. A titre d'exemple, sur la **Figure III** sont représentées les photographies prises au M.E.B., d'un trou d'air de 200 nm de diamètre et d'un réseau carré de trous d'air de 250 nm de diamètre avec une période de 690 nm.

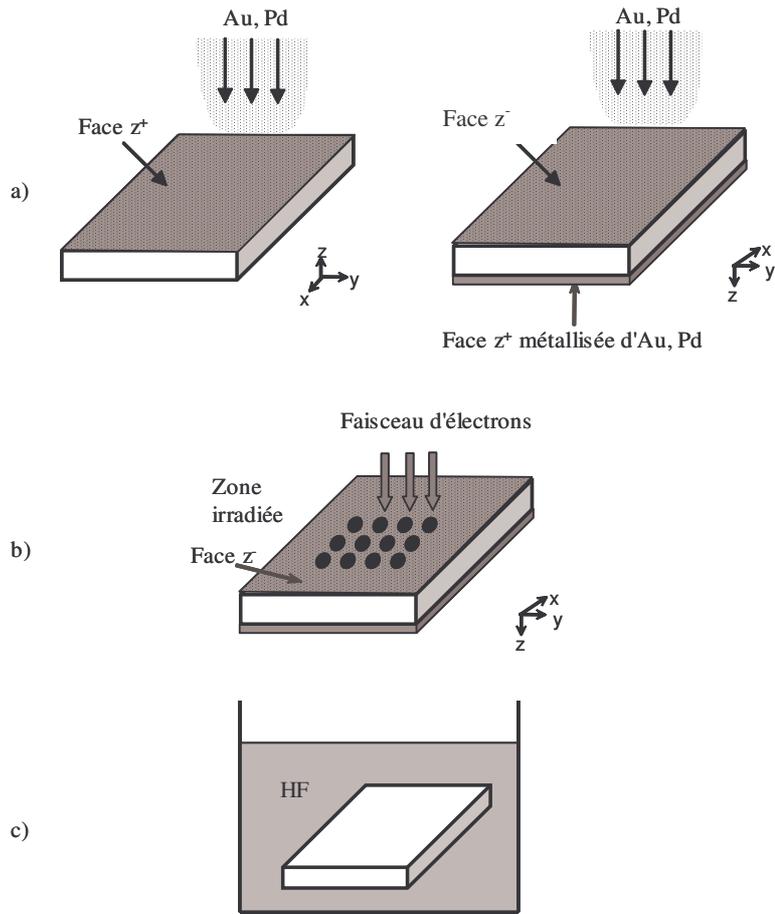
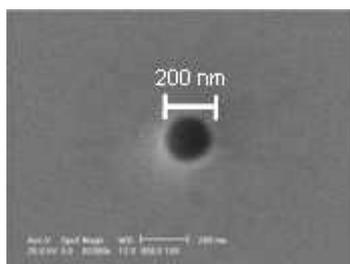
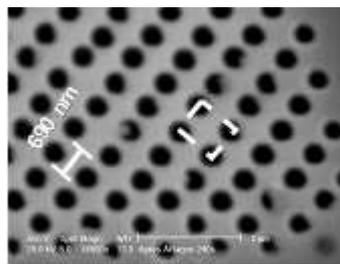


Figure II Différentes étapes du procédé expérimental utilisé pour réaliser des structures 2D



(a)



(b)

(a) trou de diamètre 200 nm
(b) réseau carré de période 690 nm, diamètre 250 nm

Figure III Photographies prises au M.E.B.

Les paramètres expérimentaux (variables d'entrée : facteurs) intervenant au cours de ces différentes étapes de fabrication ainsi que les différentes caractérisations des structures réalisées (variables de sortie : réponses) vont maintenant être présentés.

I.1.2. FACTEURS

Les différentes étapes décrites précédemment font intervenir un grand nombre de paramètres expérimentaux dont un certain nombre ont été éliminés suite à un brainstorming (grille de choix de facteurs) organisé avec l'expérimentateur et un groupe de travail. Les facteurs que nous avons choisis d'analyser sont :

- les **épaisseurs** de la couche de métallisation or/palladium. Deux épaisseurs sont à considérer : l'épaisseur de la couche sur la face z- (**Ep-**) et l'épaisseur de la couche sur la face z+ (**Ep+**),
- le **diamètre** du spot électronique,
- la **tension d'accélération** des électrons,
- la **durée totale** d'irradiation par impact,
- la **durée** d'un pulse,
- l'**intensité** du courant reçu par l'échantillon.

Les domaines de variation de ces différents paramètres sont reportés dans le **Tableau I**

Tableau I Domaines de variation des facteurs d'entrée

Facteurs principaux	Unités	Notations	Domaines de variation
épaisseur de métallisation Au/Pd face z-	nm	Ep-	2 à 50 nm
épaisseur de métallisation Au/Pd face z+	nm	Ep+	2 à 50 nm
diamètre du spot électronique	nm	spot	50 à 500 nm
tension d'accélération des électrons	kV	Tension	10 à 30 kV
durée totale d'irradiation par impact	s	t	1,2 à 240 s
durée d'un pulse	Ms	Pulse	40 à 400 ms
intensité du courant reçu par l'échantillon	nA	I	0,1 à 5 nA

I.1.3. REPONSES

Les caractéristiques géométriques et topographiques que l'on souhaite maîtriser, constituent les variables de sortie du système.

Ce sont :

- le diamètre des trous (mesuré au M.E.B.),
- la période du réseau (mesurée au M.E.B.),
- la profondeur des trous (mesurée par A.F.M. (Microscope à Force Atomique)).

I.1.4. OJECTIFS

Des simulations réalisées par l'expérimentateur, ont permis d'accéder aux dimensions requises pour obtenir des cristaux photoniques possédant une bande interdite, empêchant la propagation de la lumière à certaines longueurs d'ondes optiques. Pour obtenir une bande interdite totale à $\lambda = 1,55 \mu\text{m}$, **la période doit être égale à 397 nm** et le **diamètre des trous d'air à 336 nm**. C'est un des objectifs à atteindre. Pour cela, il est intéressant de mener des analyses de données et une démarche méthodologique associée à la construction de plans d'expériences afin :

- d'aider l'expérimentateur à obtenir de façon systématique des trous après attaque chimique,
- d'identifier les relations entre les paramètres d'entrée et de sortie c'est-à-dire de corréler le diamètre et la profondeur des trous ainsi que le pas du réseau aux facteurs d'entrée,
- d'obtenir les valeurs cibles fixées par l'expérimentateur.

Objectifs : « Obtenir un réseau de trous d'air ayant une période égale à 397 nm et un diamètre des trous d'air de 336 nm »

I.2. ANALYSES DE DONNEES

I.2.1. TABLEAU DE DONNEES

Les expériences déjà réalisées par l'expérimentateur sont répertoriées dans le tableau de données brutes (**Tableau II**).

Nous observons que des trous ne sont obtenus que pour 5 expériences. Dans un premier temps, l'objectif de cette étude sera donc de prévoir l'obtention de trous. On remarque également que les

expériences 25 et 26 apportent une information capitale sur la formation de ces trous. En effet, quand la durée passe de 80 à 180s en gardant les autres paramètres fixes, on remarque l'obtention de trous.

Nous pouvons déjà penser que ce paramètre est un paramètre influent sur l'obtention de trous. Les expériences 27 et 29 sont pour l'instant celles qui se rapprochent le plus des objectifs.

Tableau II Tableau de données brutes

Numero d'essai	Tension (kV)	t(s)	I(nA)	Ep+ (nm)	Ep- (nm)	Spot (nm)	Pulse	Résultats	Diamètres (nm)
1	10	1,6	0,51	17	17	100	400	rien	
2	12	1,6	0,59	17	17	100	400	rien	
3	15	1,6	0,64	17	17	100	400	rien	
4	20	1,6	0,68	17	17	100	400	rien	
5	10	1,6	2,2	10	10	200	400	rien	
6	12	1,6	2,28	10	10	200	400	rien	
7	15	1,6	2,85	10	10	200	400	rien	
8	20	1,6	3,21	10	10	200	400	rien	
9	25	1,6	3,73	10	10	200	400	rien	
10	30	1,6	3,93	10	10	200	400	rien	
11	10	1,6	2,47	20	20	200	400	rien	
12	12	1,6	2,47	20	20	200	400	rien	
13	15	1,6	2,47	20	20	200	400	rien	
14	20	1,6	2,47	20	20	200	400	rien	
15	25	1,6	2,48	20	20	200	400	rien	
16	30	1,6	2,48	20	20	200	400	rien	
17	25	2	3,5	20	20	200	400	rien	
18	25	4	3,5	20	20	200	400	rien	
19	25	6	3,5	20	20	200	400	rien	
20	25	8	3,5	20	20	200	400	rien	
21	25	16	3,5	20	20	200	400	rien	
22	25	20	3,5	20	20	200	400	rien	
23	25	120	3,5	20	20	200	400	rien	
24	25	40	3,5	20	20	200	400	rien	
25	25	80	3,5	20	20	200	400	rien	
26	25	180	3,5	20	20	200	400	Trou	180
27	25	240	3,5	20	20	200	400	Trou	250
28	25	0,35	3,5	20	0	200	400	Tige ?	>2000
29	25	240	3,66	20	10	200	400	Trou	410
30	25	180	3,66	20	10	200	400	Trou	200
31	10	1,2	0,045	2	2	50	40	Rien	
32	30	240	0,89	2	2	100	400	Rien	
33	10	40	0,81	2	2	200	40	Rien	
34	20	120	2,08	2	2	500	400	Rien	
35	20	40	0,51	8	8	100	40	Rien	
36	10	120	0,045	8	8	50	400	Rien	
38	10	240	0,88	8	8	200	400	Rien	
39	30	120	3,34	14	14	200	40	Rien	
40	10	40	3,8	14	14	500	400	Rien	
41	20	240	0,12	14	14	50	40	Rien	
42	10	1,2	0,2	14	14	100	400	Rien	
43	10	240	3,57	20	20	500	40	trou 700	700+ corol 1,25
44	20	1,2	2,14	20	20	200	400	Rien	
45	10	120	0,19	20	20	100	40	Tige	1330
46	30	40	0,23	20	20	50	400	Rien	
37	30	1,2	12,9	8	8	500	40	Rien	

I.2.2. BOITES A MOUSTACHES DES VARIABLES D'ENTREE

La **Figure IV** représente les boîtes à moustaches des variables d'entrée de l'étude. Nous observons que :

- la variable Intensité I possède un individu atypique (forte valeur),
- les variables Ep+, Ep-, spot et Pulse ne pourront pas être analysées car ces variables n'ont pas une variabilité homogène (très peu de valeurs différentes)

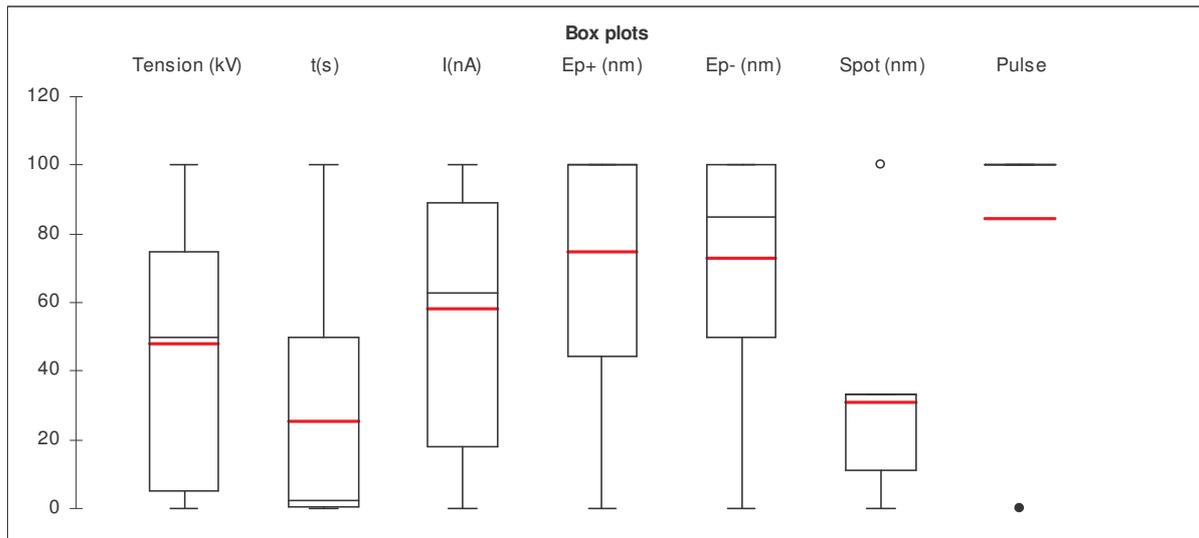


Figure IV Box plot des variables d'entrée

I.3. POTENTIELLES INTERACTIONS ET CONTRAINTES DE L'ETUDE

D'après le savoir des expérimentateurs (confirmé par les valeurs du **Tableau III**), nous savons que la tension d'accélération des électrons (**Tension**) est corrélée positivement avec l'**intensité** qui elle-même est corrélée positivement avec le diamètre du **spot**. Ceci signifie que la valeur de l'intensité (I) augmente quand l'énergie augmente. Bien que ces données ne possèdent pas une variabilité homogène nous pouvons remarquer, grâce au **Tableau III**, que le couple **Tension/spot** est corrélé à l'**intensité**.

Tableau III Tableau des coefficients de corrélation des variables d'entrée

	Tension	t	I	Ep+	Ep-	Spot	Pulse
Tension	1						
t	0,077	1					
I	0,474	-0,086	1				
Ep+	0,237	-0,067	0,129	1			
Ep-	0,125	-0,137	0,052	0,838	1		
Spot	0,050	0,075	0,669	-0,052	-0,055	1	
Pulse	0,147	-0,217	-0,034	0,331	0,250	-0,072	1

Or, il n'existe qu'un certain nombre de couples **Tension/spot** correspondant à une certaine valeur de l'intensité. Dans la méthodologie des plans d'expériences, ces facteurs sont considérés comme des facteurs « blocs ».

Il est alors fondamental de s'intéresser aux combinaisons possibles traduisant une valeur de l'énergie en fonction de ces paramètres. C'est dans ce but que **F. Quentin** a établi un tableau (**Tableau IV**) indiquant les réglages à effectuer pour obtenir des valeurs d'énergie bien définies,

de façon à pouvoir régler sans problème ces deux paramètres (plus le **Bias**) pour atteindre une énergie souhaitée.

Le **Bias** correspond au réglage de la tension de polarisation du Wehnelt (diaphragme placé immédiatement après le filament émettant les électrons), ce réglage a pour effet d'ajuster le courant d'émission du filament.

Tableau IV Etude de la valeur de l'énergie

	I Cage (nA)					
	Spot 5			Spot 6		
Tension (kV)	Bias 1	Bias 2	Bias 3	Bias 1	Bias 2	Bias 3
5				0,3	0,6	1
10	0,3	0,45	0,57	1,2	1,8	2,4
12	0,37	0,58	0,58	1,65	2,31	2,57
15	0,56	0,7		2,24	2,94	
20	0,8			3,17		
25	0,9			3,68		
30	0,9			3,5		

Ainsi, dans la méthodologie des plans d'expériences, nous pourrions considérer le facteur **Tension** comme un facteur maîtrisable.

I.4. PERSPECTIVES ET STRATEGIES ENVISAGEABLES

Dans un premier temps, il va falloir créer de l'expérimentation pour resserrer l'étude autour d'un **point de fonctionnement** c'est-à-dire déterminer les intervalles des facteurs conduisant à l'obtention de trous.

Quelles sont les possibilités de stratégies expérimentales ?

- Pour créer des données, nous pouvons recourir à la méthode des essais et des erreurs ou alors à la mise en œuvre d'une méthode d'optimisation directe (**méthode du Simplexe [2]**). L'expérimentateur pourra s'appuyer sur les réglages des expériences 27 et 29 (**Tableau V**) qui serviront de réglage de référence.

Tableau V Expériences 25 et 29

Numero d'essai	Tension (kV)	t(s)	I(nA)	Ep+ (nm)	Ep- (nm)	Spot (nm)	Pulse	Résultats	Diamètres (nm)
27	25	240	3,5	20	20	200	400	Trou	250
29	25	240	3,66	20	10	200	400	Trou	410

- Il sera utile de construire un plan de criblage pour élargir le domaine expérimental et trouver des combinaisons non testées.
- Dans tous les cas, après avoir réalisé les expériences supplémentaires conduisant à l'obtention de trous, il sera intéressant de mettre en œuvre des analyses de données, en commençant par une **Analyse Factorielle Discriminante**. Rappelons que l'**A.F.D.** est une analyse factorielle qui s'intéresse aux relations entre **une variable qualitative** (qui ici sera le résultat : obtention de trous, rien n'est observé, obtention de tiges ou obtention de trous et de tiges) et **un ensemble de variables quantitatives** qui seront ici les facteurs précédemment cités. Dans ce cas, cette analyse sera prédictive. Nous pourrions également effectuer un classement des observations nouvelles qui n'auront pas participé à la discrimination. Ces individus supplémentaires pourront être les expériences à venir d'un plan à surfaces de réponses qui nous permettra alors d'optimiser les trois réponses citées auparavant à savoir le **diamètre et la profondeur des trous** et la **période du réseau**.

Bibliographie Annexe 1

- [1] **F. Quentin, S. Massy, C. Darraud, C. Restoin, J.C. Vareille**, *Fabrication de structures 2D à l'échelle submicrométrique sur niobate de lithium par faisceau d'électrons*, 7^{èmes} Journées Nationales du Réseau Doctoral de Microélectronique (JNRDM) – Marseille – 4/7 mai 2004.
- [2] **F.H. Walters, L.R. Parker, S.L. Morgan, and S.N. Deming** *Sequential Simplex Optimization*, CRC Press, Boca Raton, USA (1991)

ANNEXE 2 : Loi normale réduite et loi de Student

1 - LOI NORMALE RÉDUITE

La variable normale réduite U suit une loi normale de moyenne 0 et d'écart-type 1 (**Figure V**).

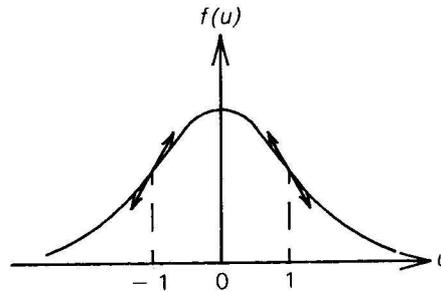


Figure V Loi normale

La table (**Tableau VI**) donne pour $u_p \geq 0$ les valeurs de la probabilité p pour que la variable centrée réduite soit inférieure à u_p , soit :

$$p = \Pr [U < u_p] = F(u_p)$$

La fonction de répartition est représentée sur la **Figure VI**.

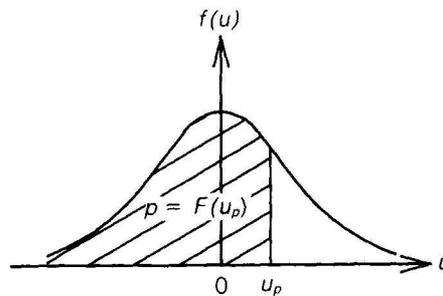


Figure VI Représentation de la fonction de répartition $F(u_p)$

Le **Tableau VI** est utilisé pour :

- donner les bornes de l'intervalle de confiance d'une moyenne au niveau $1 - \alpha$ (σ connu),
- comparer une moyenne à une valeur donnée (σ connu),
- comparer deux moyennes (σ_1 et σ_2 connus).

Tableau VI Valeurs des bornes supérieures et inférieures dans la représentation de la variable normale réduite U

	Niveau de signification	Niveau de confiance		
Intervalles de confiance et tests bilatéraux	$\alpha = 0.1$	$1-\alpha = 0.90$	$U_{1-\alpha/2} = U_{0.95}$	1.64
	$\alpha = 0.05$	$1-\alpha = 0.95$	$U_{1-\alpha/2} = U_{0.975}$	1.96
	$\alpha = 0.01$	$1-\alpha = 0.99$	$U_{1-\alpha/2} = U_{0.995}$	2.58

Soit $1 - \alpha$, le niveau de confiance attaché à un intervalle bilatéral symétrique pour U :

- la borne inférieure u_1 (**Figure VII**) est définie par : $\Pr (U < u_1) = F(u_1) = \alpha/2$. La borne inférieure s'écrit donc $u_1 = u_{\alpha/2}$.

- la borne supérieure u_2 (**Figure VII**) est telle que :

$\Pr (u < u_2) = F(u_2) = \alpha/2 + 1 - \alpha = 1 - \alpha/2$. La borne supérieure s'écrit donc $u_2 = u_{1-\alpha/2}$.

- la symétrie de la distribution permet d'écrire : $u_{\alpha/2} = - u_{1-\alpha/2}$.

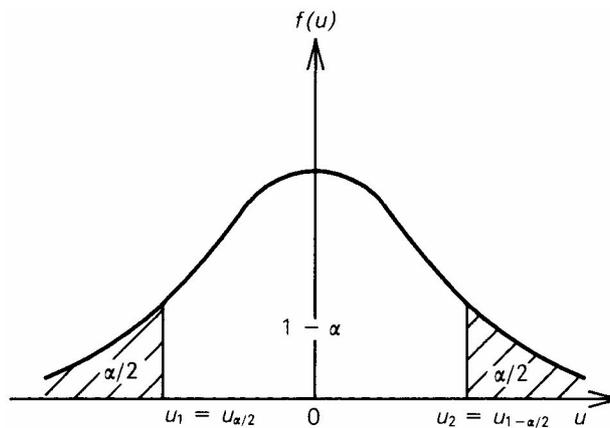


Figure VII Détermination des bornes de la courbe de la loi normale réduite

2 - LOI DE STUDENT

La variable t de Student a une distribution symétrique de moyenne nulle (**Figure VIII**) ; l'écart type dépend du nombre de degrés de liberté ν .

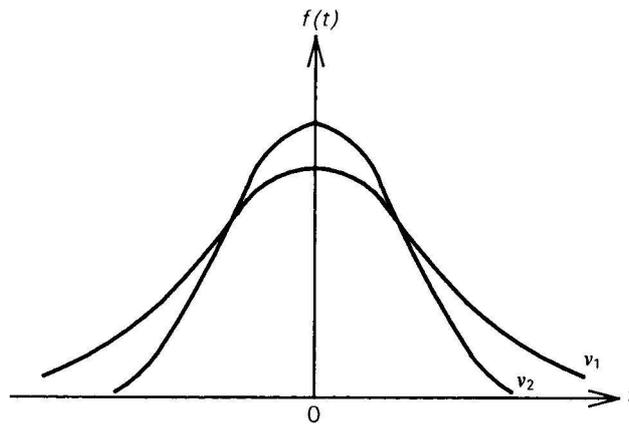


Figure VIII Représentation de la loi de Student

Quand ν tend vers l'infini, la loi de Student tend vers la loi normale réduite ; l'assimilation est acceptable à partir de $\nu = 60$. Le **Tableau VII** donne en fonction de ν les valeurs de t_p telles que :

$$\Pr (t < t_p) = F(t_p) = p \quad (p \geq 0,50)$$

La fonction de répartition est représentée sur la **Figure IX**.

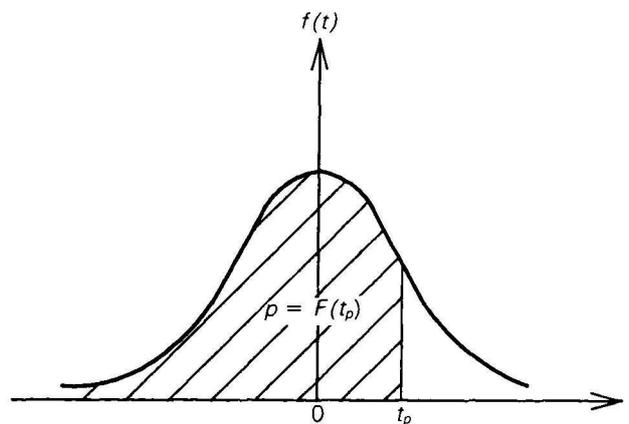


Figure IX Représentation de la fonction de répartition $F(t_p)$

Le **Tableau VII** est utilisé pour :

- donner les bornes de l'intervalle de confiance d'une moyenne au niveau $1 - \alpha$ (écart type σ estimé),
- comparer une moyenne à une valeur donnée (σ estimé),
- comparer deux moyennes (σ_1 et σ_2 égaux et estimés par s_1 et s_2).

Soit $1 - \alpha$, le niveau de confiance attaché à un intervalle bilatéral symétrique pour t . Cet intervalle a pour borne supérieure $t_{1-\alpha/2}$ et pour borne inférieure $t_{\alpha/2}$ (Figure X).

Par raison de symétrie : $t_{\alpha/2} = -t_{1-\alpha/2}$.

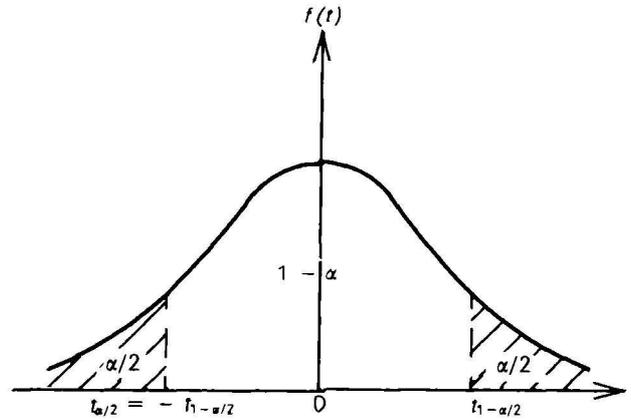


Figure X Détermination des bornes de la courbe de la loi de Student

Tableau VII Table de la loi de Student

ν \ P	0,50	0,60	0,70	0,80	0,90	0,95	0,975	0,990	0,995	0,999	0,9995
1	0,000	0,325	0,727	1,376	3,078	6,314	12,71	31,82	63,66	318,3	636,6
2	0,000	0,289	0,617	1,061	1,886	2,920	4,303	6,965	9,925	22,33	31,60
3	0,000	0,277	0,584	0,978	1,638	2,353	3,182	4,541	5,841	10,22	12,94
4	0,000	0,271	0,569	0,941	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,000	0,267	0,559	0,920	1,476	2,015	2,571	3,365	4,032	5,893	6,859
6	0,000	0,265	0,553	0,906	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,000	0,263	0,549	0,896	1,415	1,895	2,365	2,998	3,499	4,785	5,405
8	0,000	0,262	0,546	0,889	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,000	0,261	0,543	0,883	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,000	0,260	0,542	0,879	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	0,000	0,260	0,540	0,876	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,000	0,259	0,539	0,873	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	0,000	0,259	0,538	0,870	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	0,000	0,258	0,537	0,868	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	0,000	0,258	0,536	0,866	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	0,000	0,258	0,535	0,865	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	0,000	0,257	0,534	0,863	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	0,000	0,257	0,534	0,862	1,330	1,734	2,101	2,552	2,878	3,611	3,922
19	0,000	0,257	0,533	0,861	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	0,000	0,257	0,533	0,860	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	0,000	0,257	0,532	0,859	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	0,000	0,256	0,532	0,858	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	0,000	0,256	0,532	0,858	1,319	1,714	2,069	2,500	2,807	3,485	3,767
24	0,000	0,256	0,531	0,857	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	0,000	0,256	0,531	0,856	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	0,000	0,256	0,531	0,856	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	0,000	0,256	0,531	0,855	1,314	1,703	2,052	2,473	2,771	3,421	3,690
28	0,000	0,256	0,530	0,855	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	0,000	0,256	0,530	0,854	1,311	1,699	2,045	2,462	2,756	3,396	3,659
30	0,000	0,256	0,530	0,854	1,310	1,697	2,042	2,457	2,750	3,385	3,646
40	0,000	0,255	0,529	0,851	1,303	1,684	2,021	2,423	2,704	3,307	3,551
60	0,000	0,254	0,527	0,848	1,296	1,671	2,000	2,390	2,660	2,232	3,460
80	0,000	0,254	0,527	0,846	1,292	1,664	1,990	2,374	2,639	3,195	3,415
100	0,000	0,254	0,526	0,845	1,290	1,660	1,984	2,365	2,626	3,174	3,389
200	0,000	0,254	0,525	0,843	1,286	1,653	1,972	2,345	2,601	3,131	3,339
∞	0,000	0,253	0,524	0,842	1,282	1,645	1,960	2,326	2,576	3,090	3,291

(Source : CERESTA, Revue de statistique appliquée. Tables statistiques, no spécial, Paris, 1973.)

ANNEXE 3 : Loi de Snedecor

La loi de Snedecor ou loi de F permet de comparer les variances inconnues σ_1^2 et σ_2^2 (ou les écarts-types) de deux populations normales :

- σ_1^2 est estimée par s_1^2 avec ν_1 degrés de liberté,
- σ_2^2 est estimée par s_2^2 avec ν_2 degrés de liberté.

Si $\sigma_1^2 = \sigma_2^2$, le rapport $F = s_1^2/s_2^2$ suit la loi de F à ν_1 et ν_2 degrés de liberté.

La loi de F dépend de ν_1 , nombre de degrés de liberté de s_1^2 et ν_2 , nombre de degrés de liberté de s_2^2 .

Les tables (**Tableaux VIII et IX**) donnent, en fonction de ν_1 et ν_2 , les valeurs de F_p telles que :

$$\text{Prob}(F < F_p) = p \quad p \text{ peut prendre différentes valeurs : } p = 0,95 \text{ et } p = 0,975.$$

Au niveau $1 - \alpha$, les tables permettent d'obtenir :

- un intervalle de confiance bilatéral symétrique en probabilité pour F (**Figure XII**). La borne supérieure $F_{1-\alpha/2}$ est lue directement dans la table. La borne inférieure $F_{\alpha/2}$ est obtenue par la relation : $F_{\alpha/2} = 1/F_{1-\alpha/2}$ (pour ν_1 et ν_2).

- un intervalle de confiance unilatéral à droite (**Figure XIII**). Si l'éventualité $\sigma_1^2 > \sigma_2^2$ est la seule retenue, on forme le rapport $F = s_1^2/s_2^2$ que l'on compare au seuil $F_{1-\alpha}$ donné par la table.

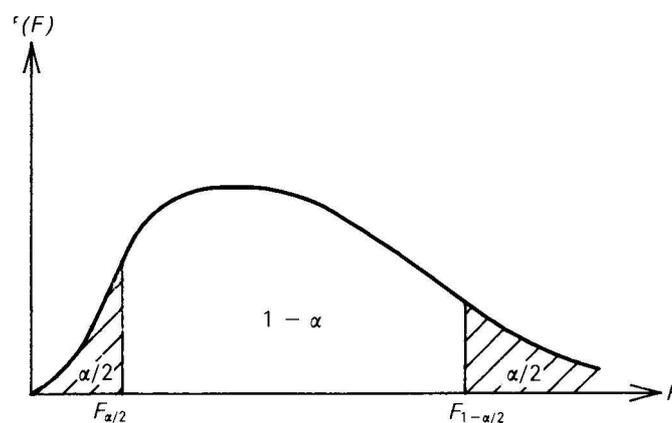


Figure XI Détermination des bornes de la courbe de la loi de Snedecor pour un intervalle de confiance bilatéral

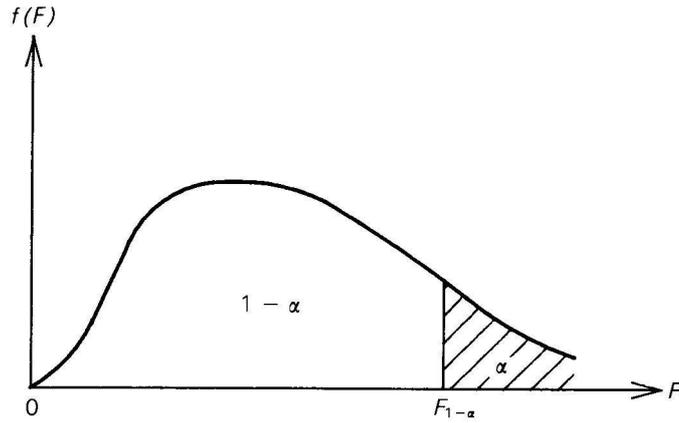


Figure XII Détermination de la borne de la courbe de la loi de Snedecor pour un intervalle de confiance unilatéral

Tableau VIII Table de la loi de Snedecor pour $p = 0,95$

$\nu_1 \backslash \nu_2$	1	2	3	4	5	6	7	8	10	12	20	40	60	100	∞
1	161	200	216	225	230	234	237	239,4	242	244	248	251	252	253	254
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	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	8,79	8,74	8,66	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	5,96	5,91	5,80	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	4,74	4,68	4,56	4,46	4,43	4,41	4,37
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,06	4,00	3,87	3,77	3,74	3,71	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,64	3,57	3,44	3,34	3,30	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,35	3,28	3,15	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	3,14	3,07	2,94	2,83	2,79	2,76	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	2,98	2,91	2,77	2,66	2,62	2,59	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,85	2,79	2,65	2,53	2,49	2,46	2,40
12	4,75	3,89	3,50	3,26	3,11	3,00	2,91	2,85	2,75	2,69	2,54	2,43	2,38	2,35	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,67	2,60	2,46	2,34	2,30	2,26	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,60	2,53	2,39	2,27	2,22	2,19	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,54	2,48	2,33	2,20	2,16	2,12	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,49	2,42	2,28	2,15	2,11	2,07	2,01
17	4,45	3,59	3,20	2,96	2,80	2,70	2,61	2,55	2,45	2,38	2,23	2,10	2,06	2,02	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,41	2,34	2,19	2,06	2,02	1,98	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,38	2,31	2,16	2,03	1,98	1,94	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,35	2,28	2,12	1,99	1,93	1,89	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,32	2,25	2,10	1,96	1,92	1,88	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,30	2,23	2,07	1,94	1,89	1,85	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,27	2,20	2,05	1,91	1,86	1,82	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,25	2,18	2,03	1,89	1,84	1,80	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,22	2,16	2,01	1,87	1,82	1,78	1,71
26	4,23	3,37	2,97	2,74	2,58	2,47	2,39	2,32	2,22	2,15	1,99	1,85	1,80	1,76	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,20	2,12	1,97	1,84	1,79	1,74	1,67
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,19	2,12	1,96	1,82	1,77	1,73	1,65
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,18	2,10	1,94	1,81	1,75	1,71	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,16	2,09	1,93	1,79	1,74	1,70	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,08	2,00	1,84	1,69	1,64	1,59	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	1,99	1,92	1,75	1,59	1,53	1,48	1,39
120	3,92	3,07	2,68	2,44	2,29	2,17	2,08	2,01	1,91	1,83	1,65	1,49	1,42	1,36	1,25
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,83	1,75	1,57	1,39	1,32	1,24	1,00

Tableau IX

Table de la loi de Snedecor pour $p = 0,975$

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	10	12	20	40	60	100	∞
1	648	800	864	900	922	937	948	957	969	977	983	1006	1010	1013	1018
2	38,5	39,0	39,2	39,2	39,3	39,3	39,4	39,4	39,4	39,4	39,4	39,5	39,5	39,5	39,5
3	17,4	16,0	15,4	15,1	14,9	14,7	14,6	14,5	14,4	14,3	14,2	14,0	14,0	14,0	13,9
4	12,2	10,6	9,98	9,60	9,36	9,20	9,07	8,98	8,84	8,75	8,56	8,41	8,36	8,32	8,26
5	10,0	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,62	6,52	6,33	6,18	6,12	6,08	6,02
6	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,46	5,37	5,17	5,01	4,96	4,92	4,85
7	8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90	4,76	4,67	4,47	4,31	4,25	4,21	4,14
8	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43	4,30	4,20	4,00	3,84	3,78	3,74	3,67
9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	3,96	3,87	3,67	3,51	3,45	3,40	3,33
10	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85	3,72	3,62	3,42	3,26	3,20	3,15	3,08
11	6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66	3,53	3,43	3,23	3,06	3,00	2,96	2,88
12	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,37	3,28	3,07	2,91	2,85	2,80	2,72
13	6,41	4,97	4,35	4,00	3,77	3,60	3,48	3,39	3,25	3,15	2,95	2,78	2,72	2,67	2,60
14	6,30	4,86	4,24	3,89	3,66	3,50	3,38	3,29	3,15	3,05	2,84	2,67	2,61	2,56	2,49
15	6,20	4,76	4,15	3,80	3,58	3,41	3,29	3,20	3,06	2,96	2,76	2,58	2,52	2,47	2,40
16	6,12	4,69	4,08	3,73	3,50	3,34	3,22	3,12	2,99	2,89	2,68	2,51	2,45	2,40	2,32
17	6,04	4,62	4,01	3,66	3,44	3,28	3,16	3,06	2,92	2,82	2,62	2,44	2,38	2,33	2,25
18	5,98	4,56	3,95	3,61	3,38	3,22	3,10	3,01	2,87	2,77	2,56	2,38	2,32	2,27	2,19
19	5,92	4,51	3,90	3,56	3,33	3,17	3,05	2,96	2,82	2,72	2,51	2,33	2,27	2,22	2,13
20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,77	2,68	2,46	2,29	2,22	2,17	2,09
21	5,83	4,42	3,82	3,48	3,25	3,09	2,97	2,87	2,73	2,64	2,42	2,25	2,18	2,13	2,04
22	5,79	4,38	3,78	3,44	3,22	3,05	2,93	2,84	2,70	2,60	2,39	2,21	2,14	2,09	2,00
23	5,75	4,35	3,75	3,41	3,18	3,02	2,90	2,81	2,67	2,57	2,36	2,18	2,11	2,06	1,97
24	5,72	4,32	3,72	3,38	3,15	2,99	2,87	2,78	2,64	2,54	2,33	2,15	2,08	2,02	1,94
25	5,69	4,29	3,69	3,35	3,13	2,97	2,85	2,75	2,61	2,51	2,30	2,12	2,05	2,00	1,91
26	5,66	4,27	3,67	3,33	3,10	2,94	2,82	2,73	2,59	2,49	2,28	2,09	2,03	1,97	1,88
27	5,63	4,24	3,65	3,31	3,08	2,92	2,80	2,71	2,57	2,47	2,25	2,07	2,00	1,94	1,85
28	5,61	4,22	3,63	3,29	3,06	2,90	2,78	2,69	2,55	2,45	2,23	2,05	1,98	1,92	1,83
29	5,59	4,20	3,61	3,27	3,04	2,88	2,76	2,67	2,53	2,43	2,21	2,03	1,96	1,90	1,81
30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,51	2,41	2,20	2,01	1,94	1,88	1,79
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,39	2,29	2,07	1,88	1,80	1,74	1,64
60	5,29	3,93	3,34	3,01	2,79	2,63	2,51	2,41	2,27	2,17	1,94	1,74	1,67	1,60	1,48
120	5,15	3,80	3,22	2,89	2,67	2,51	2,39	2,30	2,15	2,05	1,82	1,61	1,52	1,45	1,31
∞	5,02	3,68	3,12	2,79	2,57	2,41	2,29	2,19	2,05	1,94	1,71	1,48	1,39	1,30	1,00

(Source : CERESTA, Revue de statistique appliquée. Tables statistiques, n° spécial, Paris, 1973.)

Annexe 4 : Logiciels d'analyses de données et des plans d'expériences

1. XLSTAT

Ce logiciel est peu coûteux et simple d'utilisation. Il s'insère en tant que macro dans Excel. Pour l'A.C.P. et l'A.F.D, il donne tous les tableaux (les contributions, les cosinus carrés, les coordonnées des individus dans le repère de l'A.C.P.) et les trois graphiques (cercle des corrélations, projections des individus, biplot). Il est impératif de refaire l'analyse si on n'a pas choisi les bons axes. L'incorporation d'individus supplémentaires dans l'analyse n'est possible que si le nombre d'individus supplémentaires est supérieur à 1.

2. STATGRAPHICS

Ce logiciel permet l'élaboration de rapports assez simplement grâce à un traitement de texte incorporé. Il permet de faire de l'analyse de données et des plans d'expériences. Il est très utilisé pour l'analyse de données. C'est un logiciel qui n'est pas des plus intuitifs.

3. UNIWIN

Uniwin est un logiciel destiné uniquement à l'analyse multidimensionnelle de données. C'est un logiciel qui complète les analyses de Statgraphics. Il permet de réaliser des nuages de points en attribuant un poids qui détermine la grosseur du point. Il est intéressant de se servir de cet outil pour montrer la qualité de représentation des individus dans une A.C.P.

4. NEMRODW

NemrodW est un logiciel destiné essentiellement à la mise en œuvre de plans d'expériences. En effet, en ce qui concerne l'analyse de données, il ne permet que de faire de l'A.C.P. Ce logiciel est riche en choix de matrices. Il fait partie des rares logiciels permettant d'utiliser les matrices de Doelhart. Par ailleurs, il peut être difficile à utiliser pour un novice des plans d'expériences.

5. EXCEL

- Après le logiciel de traitement de texte Microsoft Word, c'est probablement un des logiciels les plus utilisés dans le monde.
- Il fait partie de l'ensemble Microsoft Office ce qui le rend accessible à un très grand nombre de personnes et de ce fait, sa compatibilité avec les autres logiciels de cet ensemble (Word, Power Point, etc.) est optimisée.
- Il s'agit d'un logiciel dont les fonctions sont très larges et peut même se prêter à une grande variété d'applications dont l'analyse statistique.

- L'utilisation de ses fonctions de base est relativement facile et son temps d'apprentissage est souvent plus court que pour la plupart des logiciels spécialisés.
- La structure de ses opérations est relativement transparente et ouverte ce qui veut dire qu'il est possible de suivre et de comprendre facilement le déroulement de ses principales opérations. De plus, il se prête facilement à des modifications puisqu'il est accompagné de divers langages de programmation.
- Son coût est relativement abordable si on le compare aux autres logiciels d'analyse statistique.
- La plupart des ordinateurs étant équipés de la suite Microsoft Office, un étudiant pourra traiter ses données de recherche dans la plupart des laboratoires ou éventuellement sur son ordinateur personnel.

Répertoire

A

A.C.P : Analyse en Composantes Principales. Cette analyse de données permet de voir les corrélations entre les variables. Deux graphiques de sortie : le cercle des corrélations et le biplot.

B

Brainstorming : technique de créativité en groupe permettant de produire le plus d'idées possibles, dans un minimum de temps sur un thème donné.

Boîte à moustaches : analyse unidimensionnelle qui permet de déceler les valeurs suspectes d'une variable.

Biplot : résultat graphique d'une A.C.P. qui regroupe sur le même schéma la projection des individus et des variables. Il permet de faire des familles d'individus.

C

Coefficients de corrélation : mesure l'intensité du lien linéaire entre x et y. La valeur du coefficient de corrélation est comprise entre -1 et 1 .

Cercle des corrélations : résultat graphique d'une A.C.P. qui permet de voir sur un cercle les différentes corrélations entre les variables.

Contribution : dans une analyse factorielle, indique la contribution des individus au calcul des axes.

Cos carré : dans une analyse factorielle, c'est un indicateur de qualité de représentation sur les nouveaux axes.

Composantes principales : nouvelles variables calculées dans une A.C.P. (z_1, z_2, \dots, z_p).

Criblage (factorial design) : plan d'expériences qui aide à dégager le poids des facteurs principaux.

Diagonalisation : opération mathématique qui permet de trouver les vecteurs propres et les valeurs propres d'une matrice.

Dendrogramme : graphique résultant d'une classification. Il regroupe les individus en un nombre restreint de classes homogènes sous forme d'arbres de classification.

E

Essais : expériences.

F

Facteur principal : variables de prédiction qui sont l'objet de l'étude et pour lesquelles on cherche à quantifier leur influence sur la réponse.

Facteur qualitatif : variables d'entrées d'un système qui prend des modalités.

Facteur quantitatif : variables d'entrées d'un système à valeur numérique continue.

I

Indice de priorité d'étude : note comprise entre 1 et 10 pour quantifier l'efficacité supposée d'un facteur sur une réponse, la faisabilité à court terme des expériences et le besoin d'information.

Individu : entité de base sur laquelle l'observateur réalise un certain nombre de mesures.

Individu supplémentaire : individu atypique dans un tableau de données. Il ne participe pas au calcul des axes dans une analyse multidimensionnelle.

Interaction : combinaison des facteurs n'agissant pas de façon indépendante.

M

Matrice d'expériences : matrice sans dimension qui permet après un recodage d'obtenir un plan d'expérimentation à réaliser.

Matrice d'expériences asymétrique : matrice d'expériences où les facteurs possèdent des nombres de modalités différentes.

Matrice d'expériences symétrique, matrice dans laquelle tous les facteurs ont le même nombre de modalités.

Matrice saturée : matrice pour laquelle le nombre d'expériences k à réaliser est égal au nombre de coefficients p du modèle.

Méthodologie des surfaces de réponses (response surface methodology) : plan d'expériences qui optimise une réponse d'un système à partir des facteurs d'entrées.

Modalité : différents états d'un facteur qui est qualitatif.

N

Nœud : nouveau point qui remplace deux individus proches dans une classification hiérarchique ascendante.

O

Orthogonalité (d'un plan d'expériences) : dans une matrice d'expériences, chaque niveau d'un facteur A est associé le même nombre de fois à chaque niveau d'un facteur B (par exemple, le niveau -1 du facteur A est associé 2 fois avec le niveau -1 et 2 fois avec le niveau 1 du facteur B).

P

Plan d'expérimentation : matrice d'expériences recodée avec les réelles valeurs des expériences à effectuer.

Plan principal : dans une A.C.P, c'est le plan formé par les axes principaux.

R

R² : carré du coefficient de corrélation multiple ou coefficient de détermination.

R_a² : coefficient de corrélation ajusté.

Randomisation : consiste à tirer au hasard l'ordre des essais pour supprimer l'influence des facteurs perturbateurs non identifiés pouvant être corrélés avec l'ordre des essais.

Réponse : d'un système (variable de sortie), critère numérique ou qualitatif dont le rôle est de traduire l'atteinte de l'objectif.

S

SCRES : somme des carrés des résidus $\sum e_i^2$.

T

Tirs numériques : résultats ou réponses d'une simulation numérique.

V

Valeurs propres : variances des individus sur les axes en analyse factorielle. Valeur diagonale de la matrice diagonale : $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_p$. Grâce à leurs valeurs numériques, on pourra calculer le pourcentage d'information porté par chaque axe.

Variable quantitative : une variable est quantitative lorsqu'elle prend ses valeurs sur une échelle numérique. C'est aussi une variable que l'on appelle continue.

Variable qualitative : une variable est qualitative lorsqu'elle prend des modalités non numériques continues. C'est aussi une variable que l'on appelle discrète.

Variable qualitative ordinale : c'est une variable qualitative obtenue à partir de notations.