



L'USINE NOUVELLE

SÉRIE | CONCEPTION

Jacques Goupy • Lee Creighton

INTRODUCTION AUX PLANS D'EXPÉRIENCES



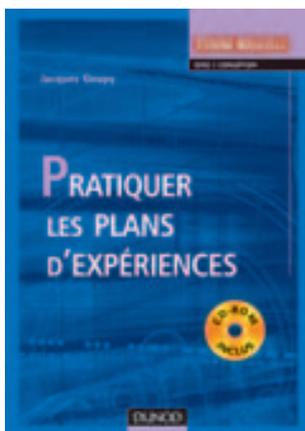
CD-Rom
offert

3^e édition

DUNOD

INTRODUCTION AUX PLANS D'EXPÉRIENCES

CHEZ LE MÊME ÉDITEUR



Jacques GOUPY
Pratiquer les plans d'expériences
(avec CD-Rom)
568 p.



Renée VEYSSEYRE
*Aide-mémoire de statistique
et probabilités pour l'ingénieur*
2^e édition, 320 p.



Jacques GOUPY
Plans d'expériences : les mélanges
304 p.

Jacques Goupy • Lee Creighton

INTRODUCTION AUX PLANS D'EXPÉRIENCES

3^e édition

L'USINE NOUVELLE

DUNOD

Du même auteur

Pratiquer les plans d'expériences (avec CD-Rom), Dunod, 2005.

Plans d'expériences : les mélanges, Dunod, 2000.

Disclaimer:

This eBook does not include the ancillary media that was packaged with the original printed version of the book.

<p>Le pictogramme qui figure ci-contre mérite une explication. Son objet est d'alerter le lecteur sur la menace que représente pour l'avenir de l'écrit, particulièrement dans le domaine de l'édition technique et universitaire, le développement massif du photocopillage.</p> <p>Le Code de la propriété intellectuelle du 1^{er} juillet 1992 interdit en effet expressément la photocopie à usage collectif sans autorisation des ayants droit. Or, cette pratique</p>	<p>d'enseignement supérieur, provoquant une baisse brutale des achats de livres et de revues, au point que la possibilité même pour les auteurs de créer des œuvres nouvelles et de les faire éditer correctement est aujourd'hui menacée.</p> <p>Nous rappelons donc que toute reproduction, partielle ou totale, de la présente publication est interdite sans autorisation de l'auteur, de son éditeur ou du Centre français d'exploitation du droit de copie (CFC, 20, rue des Grands-Augustins, 75006 Paris).</p>
--	--



© Dunod, Paris, 2001, 2006

© Bordas, Paris, 1988, pour la 1^{re} édition

ISBN 2 10 049744 8

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, 2^o et 3^o a), d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (art. L. 122-4).

Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.

TABLE DES MATIÈRES

Avant-propos	IX
1 • Faites connaissance avec la méthode des plans d'expériences	1
1.1 Processus d'acquisition des connaissances	1
1.2 Étude d'un phénomène	3
1.3 Terminologie	4
1.4 Variables centrées réduites	8
1.5 Points expérimentaux	9
1.6 Plans d'expériences	10
2 • Votre premier plan d'expériences	15
2.1 Exemple 01 : je maîtrise la consommation de ma voiture	15
3 • Plan d'expériences pour 3 facteurs	31
3.1 Exemple 02 : des bijoux en or	31
3.2 Conclusion de l'étude	47
4 • Plans factoriels complets à 4 facteurs	49
4.1 Exemple 03 : la galette des rois	50
4.2 Fonction désirabilité	60
4.3 Application à l'exemple 03	64
4.4 Conclusion de l'étude	65
5 • Notions de statistique appliquées aux plans d'expériences	67
5.1 Exemple 04 : durée de vie des outils de coupe jetables	68
5.2 Notions de statistique appliquées aux plans d'expériences	71
5.3 Facteurs et interactions influents	81
5.4 Analyse de la variance	82
5.5 Application à l'exemple 04	85
5.6 Conclusion de l'étude	91

6 • Plans factoriels fractionnaires	93
6.1 Exemple 05 : le tellure	94
6.2 Théorie des aliasés	97
6.3 Calcul de Box	99
6.4 Relation d'équivalence	102
6.5 Générateurs d'aliasés	104
6.6 Construction pratique d'un plan fractionnaire	105
6.7 Nombre maximal de facteurs étudiés sur un plan de base	112
6.8 Application de la théorie des aliasés à l'exemple 05	112
6.9 Conclusion de l'étude	113
7 • Exemples de plans factoriels fractionnaires	115
7.1 Exemple 06 : la sulfonation	116
7.2 Exemple 07 : le spectrofluorimètre	124
7.3 Exemple 08 : les pommes de terre chips	140
8 • Ordre des essais	147
8.1 Nature des erreurs	148
8.2 Exemple 09 : le <i>Penicillium chrysogenum</i> (exemple de blocking)	151
8.3 Exemple 10 : les haricots de Yates	161
8.4 Exemple 11 : le broyeur (exemple de plan anti-dérive)	168
8.5 Avantages et dangers de la randomisation	175
9 • Plans pour surfaces de réponse	179
9.1 Présentation des plans composites	179
9.2 Présentation des plans de Box-Behnken	180
9.3 Présentation des plans de Doehlert	181
9.4 Exemple 12 : la rectification du contremaître (exemple de plan composite)	184
9.5 Exemple 13 : un yoghourt doux (exemple de plan de Box-Behnken)	194
9.6 Exemple 14 : l'insecticide (exemple de plan de Doehlert)	199
10 • Plans de mélanges	207
10.1 Contrainte fondamentale des mélanges	207
10.2 Représentation géométrique des mélanges	208
10.3 Plans de mélanges classiques	212
10.4 Modèles mathématiques des plans de mélanges	215
10.5 Exemple 15 : les trois polymères	217

11 • Notion de plan optimal	221
11.1 Exemple d'Hotelling	221
11.2 Pesées et plans d'expériences	223
11.3 Optimalité	225
11.4 Plans optimaux avec un modèle linéaire	229
11.5 Quand utiliser des plans optimaux ?	232
11.6 Adaptabilité des plans optimaux	232
11.7 Exemple 16 : développement d'un détecteur de fissures	235
11.8 Exemple 17 : les comprimés du pharmacien	240
11.9 Exemple 18 : la crème à bronzer	248
12 • Carrés latins et plans associés	255
12.1 Exemple 19 : les salaires – un facteur discret à plusieurs niveaux	255
12.2 Suite de l'exemple 19 – deux facteurs discrets à plusieurs niveaux	257
12.3 Carrés latins	264
12.4 Carrés gréco-latins	265
12.5 Carrés de Youden	266
12.6 Exemple 20 : étude de la pénétrométrie	267
13 • Synthèse et conseils	271
13.1 Choix d'une méthode d'expérimentation	271
13.2 Expérimentation	277
13.3 Analyse des résultats	277
13.4 Acquisition progressive des connaissances	279
13.5 Recommandations	280
13.6 Limites de l'expérimentique	280
14 • Préparez et analysez vos plans d'expériences avec JMP®	281
14.1 Construire un plan d'expériences	281
14.2 Le tableur de JMP	286
14.3 Choisir et utiliser un modèle	288
14.4 Examiner et analyser les résultats	290
14.5 Principaux graphiques et rapports d'analyse	291
14.6 Transférer et sauvegarder les résultats et les analyses	293

Annexes

A • Origine des exemples	297
B • Comparaison de deux moyennes indépendantes	299
C • Séries d'essais laissant les facteurs principaux insensibles à une dérive linéaire	301
D • Introduction au calcul matriciel	307
Bibliographie	313
Index	321
Instructions pour utiliser le logiciel JMP®	324

AVANT-PROPOS

La première édition de ce livre a été publiée en 1988. Depuis cette année-là, l'utilisation des plans d'expériences s'est considérablement développée dans les mondes industriels et universitaires. De nombreux logiciels consacrés aux plans d'expériences sont apparus. Ils ont simplifié la construction des plans, effacé les difficultés de calcul et permis la réalisation rapide de graphiques mettant en valeur les principaux résultats de l'expérimentation. Il était donc devenu nécessaire de rajeunir l'ouvrage *Introduction aux plans d'expériences* pour tenir compte de cette rapide évolution. C'est l'objet de cette troisième édition.

Les première et deuxième éditions ne traitaient que des plans factoriels, complets ou fractionnaires. Cette troisième édition comporte, en plus, des chapitres sur les plans pour surfaces de réponse, sur les plans de mélanges, sur les plans D-optimaux et sur les plans pour variables discrètes. Cet élargissement a été possible grâce aux logiciels d'aujourd'hui qui permettent de construire facilement tous ces plans et qui réalisent les calculs les plus complexes rapidement et sans aucune difficulté.

Pour que vous puissiez vous rendre compte de la puissance des logiciels actuels, la société SAS nous a permis d'incorporer une version complète de son logiciel JMP® (prononcez *jump*) dans ce livre. Vous aurez ce logiciel à votre disposition pendant six mois à partir du moment où vous l'aurez activé (voir les instructions en fin d'ouvrage).

Ce livre est complété par un blog (<http://www.plansdexperiences.com>) sur lequel vous trouverez des explications supplémentaires sur les exemples, un manuel (en français) d'introduction au logiciel et les fichiers téléchargeables des exemples. Une documentation plus complète existe en anglais et peut être consultée et téléchargée sur le site <http://www.jmp.com>.

Nous tenons à remercier Michel Genetay, Rémi Longerey, Lalatiana Rakotozafy et Jacques Vaillé qui ont relu le manuscrit original de ce livre et qui nous ont fait profiter de leurs suggestions et de leurs conseils.

Nous conseillons aux lecteurs qui veulent compléter leurs connaissances, d'accompagner leur lecture par l'exécution des calculs avec le logiciel. Ils pourront ainsi approfondir les analyses et procéder à leur propre interprétation. Nous avons essayé de fournir aux lecteurs tous les outils et toutes les explications pour qu'ils puissent utiliser facilement la méthode des plans d'expériences pour leurs propres travaux. Il nous reste à leur souhaiter : « Bons plans d'expériences ! »

Jacques Goupy, Paris (France)

Lee Creighton, Cary (Caroline du Nord, États-Unis)

1 • FAITES CONNAISSANCE AVEC LA MÉTHODE DES PLANS D'EXPÉRIENCES

Si vous lisez ce livre c'est que, probablement, vous faites des expériences et que vous cherchez à mieux les organiser. Vous cherchez surtout à ne faire que les bonnes expériences, c'est-à-dire celles qui vont vous apporter la meilleure information expérimentale que vous puissiez espérer.

Ce livre est écrit pour vous aider. Il va vous décrire la méthode et les outils pour que vous ne fassiez plus que les bonnes expériences et que vous en tiriez le maximum de renseignements. Vous prendrez vos décisions en connaissance de cause et vous trouverez des solutions à vos problèmes.

Nous allons d'abord regarder comment s'intègrent les plans d'expériences dans le processus d'acquisition des connaissances. Puis, nous verrons les concepts de base qui permettent de bien définir une étude et comment l'on interprète les résultats des essais.

1.1 Processus d'acquisition des connaissances

Augmenter ses connaissances, c'est trouver la réponse à une question posée. On commence donc par se poser une ou plusieurs questions (Figure 1.1). Par exemple, si l'on désire connaître l'influence d'un engrais sur le rendement en blé d'une parcelle de terrain, on pourra se poser les questions suivantes qui, bien sûr, ne sont pas limitatives :

- « Peut-on obtenir 10 quintaux de blé de plus à l'hectare en augmentant la quantité d'engrais ? »
- « Quelle est l'incidence de la pluie sur l'efficacité de l'engrais ? »
- « La qualité du blé restera-t-elle bonne si j'utilise cet engrais ? »
- « Combien dois-je utiliser d'engrais pour obtenir la quantité maximale de blé par hectare ? »

Ces questions délimitent le problème à résoudre et fixent les travaux à exécuter. Il est donc important de se poser les questions qui correspondent réellement au problème. Bien sûr, avant d'entreprendre des expériences, il est prudent de vérifier que l'information cherchée n'existe pas déjà ailleurs. Dans ce but, on effectue un inventaire des connaissances, soit une bibliographie, soit une consultation d'experts, soit un calcul théorique, soit toute autre démarche ayant pour but de répondre,

sans expérience, aux questions posées. Après cette phase d'enquête, les questions initiales peuvent être entièrement résolues et il n'y a plus de problème. Dans le cas contraire, certaines questions restent entières ou elles sont modifiées : il est alors nécessaire d'entreprendre des expériences pour obtenir les réponses complètes que l'on souhaite. Ce travail préalable fait partie du métier d'expérimentateur et nous n'insisterons pas. Ce n'est pas sur cette phase du processus que nous voulons faire porter notre effort de réflexion et d'amélioration.

N'ayant plus que les questions sans réponse immédiate, il faut faire des expériences. Comment peut-on choisir ces expériences pour :

- arriver rapidement aux meilleurs résultats possibles,
- éviter de réaliser des expériences inutiles,
- obtenir la meilleure précision possible sur les résultats,
- permettre d'avancer à coup sûr,
- établir la modélisation du phénomène étudié,
- découvrir la solution optimale.

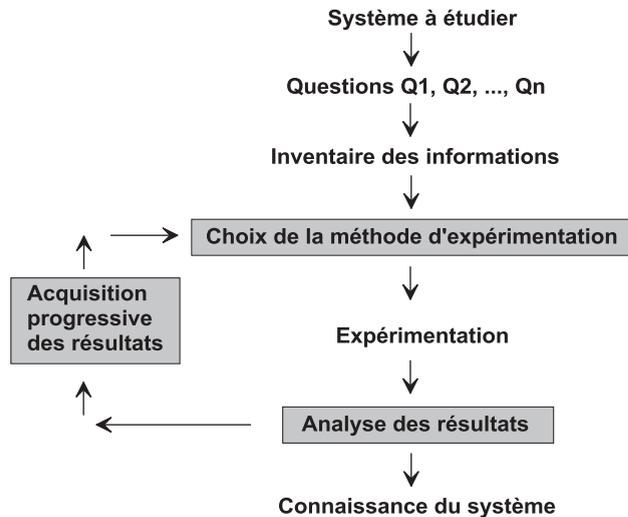


Figure 1.1 – Les plans d'expériences optimisent les trois parties encadrées du processus d'acquisition des connaissances.

Les trois aspects essentiels du processus d'acquisition des connaissances sont les suivantes :

- le choix de la méthode d'expérimentation,
- l'analyse des résultats,
- l'acquisition progressive de la connaissance.

Examinons plus en détail ces trois aspects sachant que les expériences sont organisées pour faciliter l'exploitation des résultats et pour permettre l'acquisition progressive des résultats d'intérêt.

1.1.1 Choix de la méthode d'expérimentation

La méthode d'expérimentation choisie doit faciliter l'interprétation des résultats. Elle doit également minimiser le nombre des essais sans toutefois sacrifier la qualité. La théorie des plans d'expériences assure les conditions pour lesquelles on obtient la meilleure précision possible avec le minimum d'essais. On a donc le maximum d'efficacité avec le minimum d'expériences et par conséquent le coût minimum.

1.1.2 Analyse des résultats

L'analyse des résultats d'expériences est facilitée par le choix initial des expériences. Les résultats seront faciles à interpréter et riches d'enseignement si l'on a bien préparé les expériences.

Grâce aux ordinateurs et aux logiciels la construction des plans d'expériences et les calculs nécessaires à leur interprétation sont devenus très simples. Ces outils favorisent également les représentations graphiques qui illustrent de manière spectaculaire les résultats et améliorent la compréhension des phénomènes.

1.1.3 Acquisition progressive de la connaissance

L'expérimentateur qui entreprend une étude ne connaît pas les résultats, il est donc sage d'avancer progressivement pour pouvoir réorienter les essais en fonction des premiers résultats. Une première ébauche permettra de mieux orienter les essais vers les seuls aspects intéressants de l'étude et d'abandonner les voies sans issues.

Une première série d'expériences conduit à des conclusions provisoires ; en fonction de ces conclusions provisoires, on lance une nouvelle série d'essais. L'ensemble des deux séries d'expériences est utilisé pour obtenir un contour précis des résultats d'étude. L'expérimentateur accumule ainsi les seuls résultats dont il a besoin et s'arrête dès qu'il a obtenu ce qu'il cherche.

1.2 Étude d'un phénomène

L'étude d'un phénomène revient souvent à s'intéresser à une grandeur particulière comme la consommation d'essence d'une voiture ou comme le prix de revient d'un produit chimique ou encore comme le rendement en blé d'une parcelle de terrain. Cette grandeur, consommation, prix ou rendement, dépend d'un grand nombre de variables. La consommation de la voiture est fonction de la vitesse du véhicule, de la puissance du moteur, de la manière de conduire, de la direction et de la force du vent, du gonflage des pneumatiques, de la présence ou non d'un porte-bagages, du nombre de personnes transportées, de la marque de la voiture, etc. Le prix du produit chimique dépend de la qualité des matières premières, des rendements des unités de production, des spécifications imposées, des conditions de fabrication, etc. Il en est de même pour le rendement en blé qui est fonction de la nature du terrain, de la quantité d'engrais incorporé, de l'exposition au soleil, du climat, de la variété de blé semencé, etc.

Sous une forme mathématique, on peut écrire que la grandeur d'intérêt, y , que nous appellerons également *réponse* par la suite, est une fonction de plusieurs variables x_i (variables que nous appellerons aussi *facteurs* par la suite). On a :

$$y = f(x_1, x_2, x_3, \dots, x_k)$$

L'étude du phénomène se ramène à déterminer la fonction $f()$ qui lie la réponse y aux différents facteurs x_1, x_2, \dots, x_k .

Pour approfondir cette approche il faut introduire quelques notions particulières et une terminologie spécifique aux plans d'expériences.

1.3 Terminologie

La grandeur d'intérêt, qui est généralement notée y , porte le nom de *réponse*. Les variables qui peuvent modifier la réponse sont appelées *facteurs*. On parle donc des facteurs qui influent sur une réponse. Les termes *facteur* et *réponse* sont universellement employés dans le domaine des plans d'expériences.

1.3.1 Les différents types de facteurs

La construction des plans et l'interprétation des résultats dépendent en grande partie des types de facteurs rencontrés dans l'étude. On distingue plusieurs types de facteurs. Nous retiendrons les types de facteurs suivants : les facteurs continus, les facteurs discrets, les facteurs ordonnables, les facteurs booléens.

■ Facteurs continus

La pression est un exemple de facteur continu. Dans un intervalle de pression donné, on peut choisir toutes les valeurs possibles. Il en est de même d'une longueur, d'une concentration ou d'une température. Les valeurs prises par les facteurs continus sont donc représentées par des nombres continus.

■ Facteurs discrets

Au contraire, les facteurs discrets ne peuvent prendre que des valeurs particulières. Ces valeurs ne sont pas forcément numériques : on peut représenter un facteur discret par un nom, une lettre, une propriété ou même par un nombre qui n'a alors en soi aucune valeur numérique mais qu'une signification de repère. Par exemple, on peut s'intéresser aux couleurs d'un produit : bleu, rouge et jaune sont des facteurs discrets.

■ Facteurs ordonnables

Il s'agit de facteurs discrets que l'on peut mettre dans un ordre logique. Par exemple, grand, moyen, petit, ou encore premier, deuxième, troisième et quatrième.

■ Facteurs booléens

Les facteurs booléens sont des facteurs discrets qui ne peuvent prendre que deux valeurs : haut ou bas, ouvert ou fermé, blanc ou noir, etc.

La frontière est parfois floue entre ces différents types de facteurs. Une couleur, facteur apparemment discret, peut être transformée en grandeur ordonnable et même continue si l'on introduit la notion de longueur d'onde. Un facteur continu, comme la vitesse, peut être transformé en facteur ordonnable ou discret : rapide et lent, ou vitesse A et vitesse B. Cette possibilité n'est pas un inconvénient, c'est une souplesse supplémentaire dont l'expérimentateur pourra tirer avantage au moment de l'interprétation. En effet, suivant le choix effectué, on pourra mettre plus facilement en relief un aspect particulier de l'étude. Le changement de types de variables est aussi un moyen d'adapter la réponse à l'objectif de l'étude. Par exemple, on possède l'âge des membres d'une population. Si l'on cherche l'âge moyen, la variable « âge » sera choisie continue. Si l'on cherche le pourcentage des personnes ayant un âge donné, la variable « âge » sera choisie ordonnable : il y aura plusieurs catégories définies par les limites d'âge, les jeunes, les adolescents, les adultes et les seniors. Si l'on cherche la proportion de jeunes de moins de 18 ans, la variable « âge » sera choisie booléenne : les moins de 18 ans et les plus de 18 ans. La compréhension des plans d'expériences s'appuie sur la notion essentielle d'espace expérimental des grandeurs étudiées. Nous allons examiner en détail ce concept fondamental. Nous insisterons sur les facteurs continus qui sont les plus couramment utilisés.

1.3.2 Espace expérimental

Pour présenter l'espace expérimental nous utiliserons un espace à deux dimensions, ce qui facilitera les représentations graphiques. Il est ensuite facile d'étendre les notions introduites à des espaces multidimensionnels.

Un facteur continu peut être représenté par un axe gradué et orienté. S'il y a un second facteur continu, il est représenté, lui aussi, par un axe gradué et orienté. Ce second axe est disposé orthogonalement au premier. On obtient ainsi un repère cartésien qui définit un espace euclidien à deux dimensions. Cet espace est appelé *l'espace expérimental* (figure 1.2). L'espace expérimental comprend tous les points du plan « facteur 1 \times facteur 2 » et chacun d'eux représente une expérience.

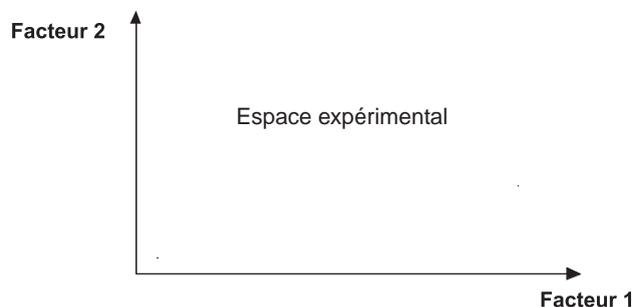


Figure 1.2 – Chaque facteur est représenté par un axe gradué et orienté. Les axes des facteurs sont orthogonaux entre eux. L'espace ainsi défini est l'espace expérimental.

S'il y a un troisième facteur, on le représente aussi par un axe orienté et gradué, et on le positionne perpendiculairement aux deux premiers. À partir de quatre facteurs, on opère de même, mais il n'y a plus de représentation géométrique possible et l'on doit adopter une représentation purement mathématique de l'espace expérimental qui est un hypervolume à quatre dimensions.

1.3.3 Domaine d'un facteur

La valeur donnée à un facteur pour réaliser une expérience est appelée *niveau*.

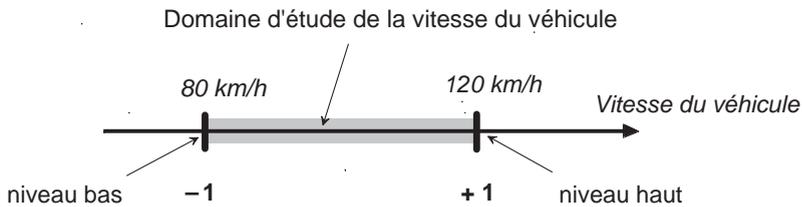


Figure 1.3 – Le domaine de variation de la vitesse est constitué de toutes les vitesses comprises entre 80 et 120 km/h. Le niveau bas du facteur est noté -1 et le niveau haut $+1$.

Lorsqu'on étudie l'influence d'un facteur, en général, on limite ses variations entre deux bornes. La borne inférieure est le *niveau bas*. La borne supérieure est le *niveau haut*. Si l'on étudie l'influence de la vitesse du véhicule sur la consommation, celle-ci peut varier, par exemple, entre 80 et 120 km/h. La vitesse de 80 km/h est le niveau bas et la vitesse de 120 km/h est le niveau haut. C'est l'expérimentateur qui définit ces deux niveaux en fonction des spécificités de l'étude. L'ensemble de toutes les valeurs que peut prendre le facteur entre le niveau bas et le niveau haut, s'appelle le *domaine de variation du facteur* ou plus simplement le *domaine du facteur*.

S'il y a plusieurs facteurs, chacun d'eux à son domaine de variation. Afin d'avoir une représentation commune pour tous les facteurs, on a l'habitude d'indiquer les niveaux bas par -1 et les niveaux hauts par $+1$. La vitesse de 80 km/h est le niveau -1 et celle de 120 km/h est le niveau $+1$.

À l'intérieur du domaine d'un facteur continu toutes les valeurs sont théoriquement possibles. On peut donc y choisir deux, trois ou plusieurs niveaux selon les besoins de l'étude. Par exemple, si l'on veut établir un modèle du second degré, on choisira trois ou quatre niveaux, c'est-à-dire trois ou quatre vitesses différentes.

1.3.4 Domaine d'étude

Dans la pratique, l'expérimentateur sélectionne une partie de l'espace expérimental pour réaliser son étude. Cette zone particulière de l'espace expérimental est le *domaine d'étude* (figure 1.4). Ce domaine est défini par les niveaux hauts et les niveaux bas de tous les facteurs et éventuellement par des contraintes entre les facteurs. Supposons que le second facteur soit la surcharge du véhicule définie comme toute masse supplémentaire à celle du véhicule et du chauffeur. Le niveau

bas de la surcharge est 0 kg et le niveau haut 300 kg, par exemple. S'il n'y a pas de contraintes, le domaine d'étude est représenté par tous les points dont les surcharges sont comprises entre 0 et 300 kg et dont les vitesses sont comprises entre 80 et 120 km/h.

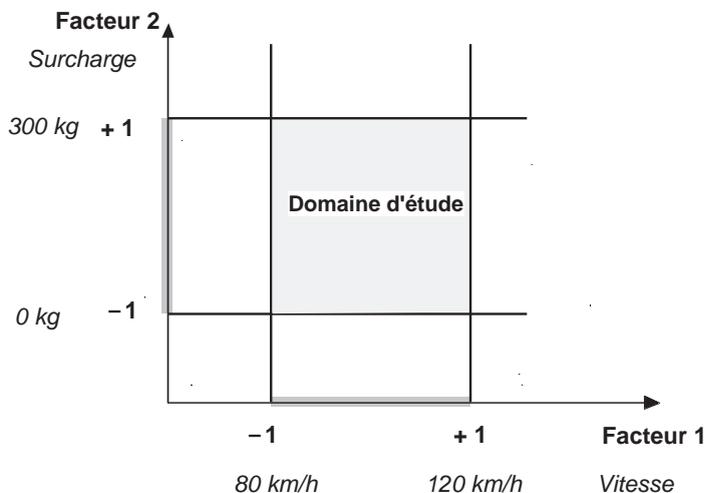


Figure 1.4 – Le domaine d'étude est défini par la réunion des domaines des différents facteurs (ici, il n'y a pas de contraintes).

Il peut y avoir des contraintes sur le domaine d'étude. Par exemple, il peut être impossible d'atteindre la vitesse de 120 km/h avec une surcharge trop élevée. La figure 1.5 illustre une réduction possible du domaine d'étude initial. Une zone du domaine d'étude initial échappe aux expériences.

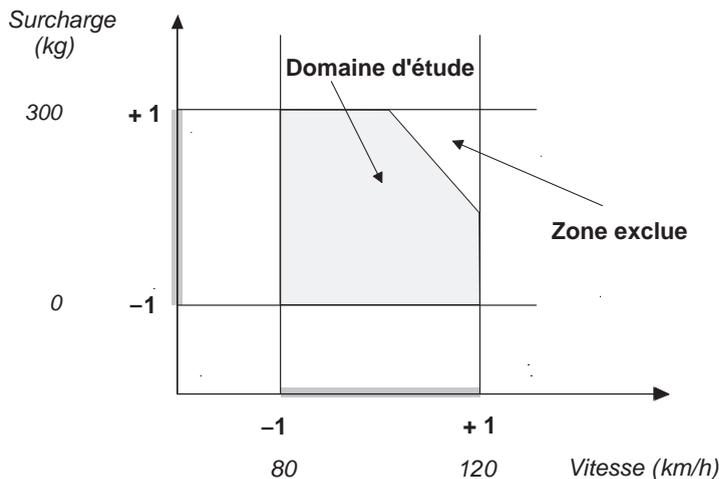


Figure 1.5 – Le domaine d'étude sous contraintes est représenté par la partie grisée.

1.4 Variables centrées réduites

Lorsqu'on attribue la valeur -1 au niveau bas d'un facteur et la valeur $+1$ au niveau haut, on effectue deux modifications importantes :

- On déplace l'origine des mesures. Dans l'exemple choisi, le milieu de l'intervalle $[-1 ; +1]$ correspond à une valeur de 100 km/h. La valeur numérique de la nouvelle origine, zéro, diffère donc de l'origine exprimée en unité courante.
- On change l'unité de mesure. Par exemple, si le niveau bas du facteur « vitesse du véhicule » est 80 km/h et le niveau haut 120 km/h, il y a 40 km/h entre ces deux valeurs, soit 40 fois l'unité de vitesse. Entre -1 et $+1$ il y a deux unités nouvelles : la nouvelle unité vaut 20 km/h, on lui donne le nom de *Pas*.

Ces deux modifications entraînent l'introduction de nouvelles variables que l'on appelle *variables centrées réduites* (v.c.r) : centrées pour indiquer le changement d'origine et réduites pour signaler la nouvelle unité. On utilise également le terme de variables codées ou d'unités codées.

Le passage des variables d'origine A aux variables codées x , et inversement, est donné par la formule suivante (A_0 est la valeur centrale en unités courantes) :

$$x = \frac{A - A_0}{Pas} \quad \{1.1\}$$

L'intérêt des unités codées est de pouvoir présenter les plans d'expériences de la même manière quels que soient les domaines d'étude retenus et quels que soient les facteurs. La théorie des plans d'expériences présente ainsi une grande généralité.

L'utilisation des v.c.r est très répandue dans les logiciels de plans d'expériences et certaines opérations comme la recherche des meilleurs points d'expériences par le critère de D-optimalité ne sont réalisables qu'avec ces variables.

Les variables codées résultent du rapport de deux grandeurs de même unité physique, elles sont donc sans dimension. La disparition des unités naturelles associée au fait que tous les facteurs ont le même domaine de variation (deux unités codées) permet la comparaison directe des effets des facteurs entre eux.

Application 1

Un expérimentateur choisit pour le facteur « vitesse de la voiture » 80 km/h comme niveau bas et 120 km/h comme niveau haut. Quelle est, en variables codées, la valeur de la vitesse de 90 km/h ?

Calculons le *Pas* du facteur vitesse. Il est égal à la moitié de la différence entre le niveau haut et le niveau bas :

$$Pas = \frac{A_{+1} - A_{-1}}{2} = \frac{120 - 80}{2} = 20$$

A_0 est la valeur centrale entre le niveau haut et le niveau bas, c'est-à-dire la moitié de la somme du niveau haut et du niveau bas :

$$A_0 = \frac{A_{+1} + A_{-1}}{2} = \frac{120 + 80}{2} = 100$$

Appliquons la relation {1.1} :

$$x = \frac{A - A_0}{Pas} = \frac{90 - 100}{20} = -0,5$$

Pour cet exemple, la vitesse de 90 km/h est donc égale à $-0,5$ en variables codées.

Application 2

On peut aussi chercher la valeur normale connaissant la valeur codée. Quelle est, en unités normales, la valeur de la vitesse de +0,5 en unités codées ? Écrivons la relation {1.1} :

$$+0,5 = \frac{A - 100}{20}$$

D'où : $A = +100 + 0,5 \times 20 = 110$

La vitesse codée de 0,5 correspond à une vitesse de 110 km/h.

1.5 Points expérimentaux

Dans un espace à deux dimensions, le niveau i du facteur 1, noté $x_{1,i}$, et le niveau j du facteur 2, noté $x_{2,j}$, peuvent être considérés comme les coordonnées d'un point de l'espace expérimental ou du domaine d'étude (figure 1.6). Par exemple, si le niveau de la vitesse est 90 km/h et celui de la surcharge est 250 kg, les coordonnées du point expérimental sont :

$$\begin{aligned}x_{1,i} &= 90 \text{ km/h} \\x_{2,j} &= 250 \text{ kg}\end{aligned}$$

Une expérience donnée est alors représentée par un point dans ce système d'axes. C'est la raison pour laquelle une expérience est souvent désignée par l'expression *point expérimental*, *point d'expérience* ou même simplement *point*. Un plan d'expériences est donc représenté par un ensemble de points expérimentaux, eux-mêmes situés dans l'espace expérimental. Dans l'exemple que nous avons pris, l'expérience est conduite sur un véhicule qui roule à 90 km/h avec une surcharge de 250 kg.

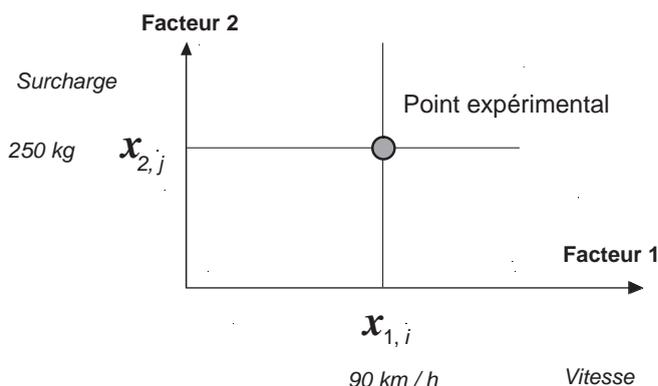


Figure 1.6 – Dans l'espace expérimental, les niveaux des facteurs définissent des points expérimentaux.

Jusqu'à trois facteurs, il est possible de dessiner le domaine d'étude. Au-delà de trois facteurs, on utilise une représentation en tableau, dite *matricielle*, plus générale puisqu'elle permet de représenter les points d'expériences dans un hypervolume à un nombre quelconque de dimensions.

1.6 Plans d'expériences

1.6.1 Méthodologie des plans sans contrainte

Le choix du nombre et de l'emplacement des points d'expériences est le problème fondamental des plans d'expériences. On cherche à réaliser le minimum d'expériences tout en réduisant l'influence de l'erreur expérimentale sur les modélisations mathématiques qui serviront à prendre des décisions. On atteint ce but en considérant les propriétés mathématiques et statistiques qui relient la réponse aux facteurs. Lorsqu'il n'y a pas de contraintes sur le domaine d'étude, il existe des plans classiques qui possèdent d'excellentes qualités statistiques et qui permettent de modéliser les réponses dans les meilleures conditions (figure 1.7). Lorsqu'il existe des contraintes, il faut construire des plans sur mesure en recherchant la position des points expérimentaux qui conduisent, là aussi, à de bonnes qualités statistiques et à une bonne modélisation des réponses.

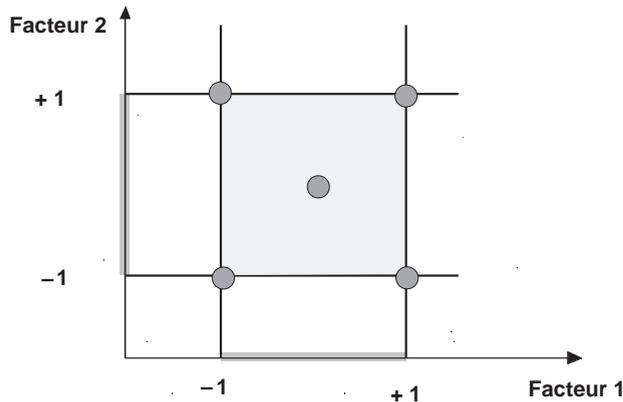


Figure 1.7 – Exemple de disposition des points expérimentaux dans un domaine sans contraintes.

1.6.2 Méthodologie des plans avec contrainte

La procédure de construction des plans dont le domaine est contraint est la suivante :

1. On définit le domaine de chacun des facteurs (niveau bas et niveau haut).
2. On définit les contraintes qui pèsent sur les facteurs. Ces contraintes sont exprimées par des relations d'inégalité entre les facteurs et elles définissent les zones autorisées, c'est-à-dire celles où les expériences sont possibles, et les zones interdites, c'est-à-dire celles où les expériences ne doivent pas être exécutées.
3. On définit les niveaux des facteurs les plus intéressants pour l'étude, autres que les niveaux bas et hauts. Le plus souvent entre 2 et 5 niveaux supplémentaires par facteur.
4. On construit une grille en prenant en compte toutes les combinaisons des niveaux des facteurs. Cette grille ne doit contenir que les points expérimentaux réalistes, c'est-à-dire les points du domaine autorisé. Ces points constituent les points candidats (figure 1.8).

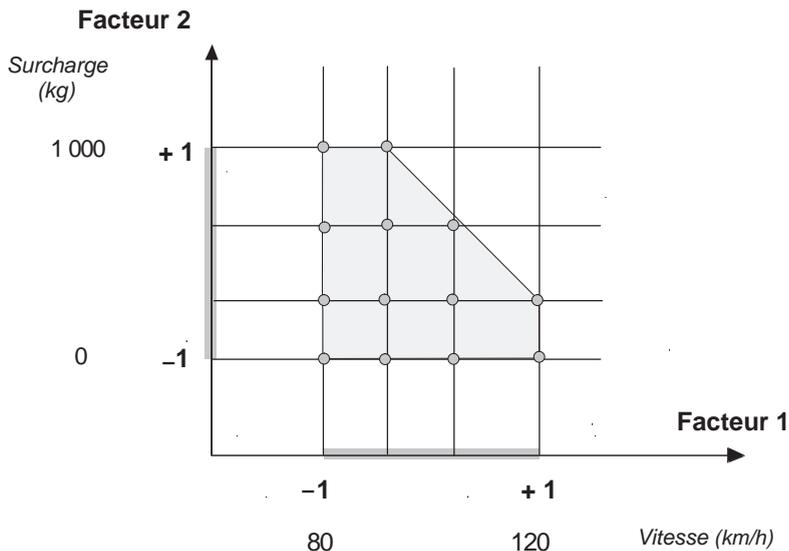


Figure 1.8 – La grille des points candidats est constituée d'expériences possibles dans le domaine d'étude.

5. On choisit une fonction reliant *a priori* la réponse aux facteurs.

6. On sélectionne, en fonction d'un critère d'optimalité choisi, le nombre et l'emplacement des points expérimentaux les plus utiles à la modélisation du phénomène étudié (figure 1.9). Cette sélection exige de longs calculs et n'est possible qu'avec l'aide d'un logiciel de plans d'expériences.

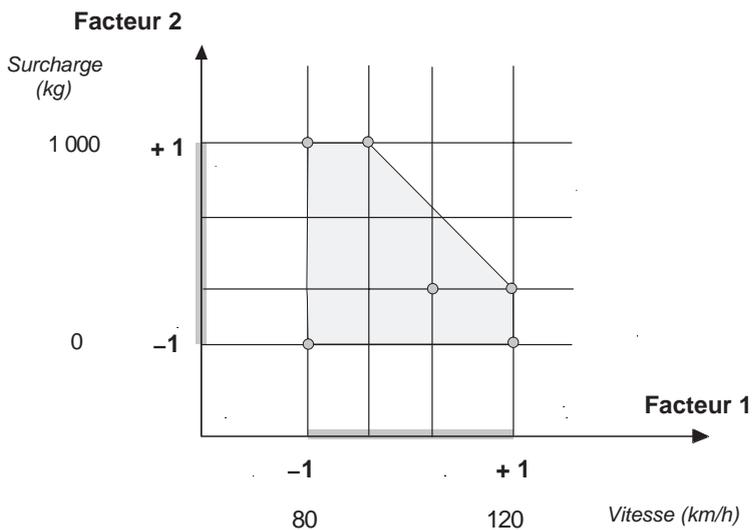


Figure 1.9 – Les meilleurs points sont sélectionnés par le logiciel.

1.6.3 Surfaces de réponse

À chaque point du domaine d'étude correspond une réponse. À l'ensemble de tous les points du domaine d'étude correspond un ensemble de réponses qui se localise sur une surface appelée *surface de réponse* (figure 1.10).

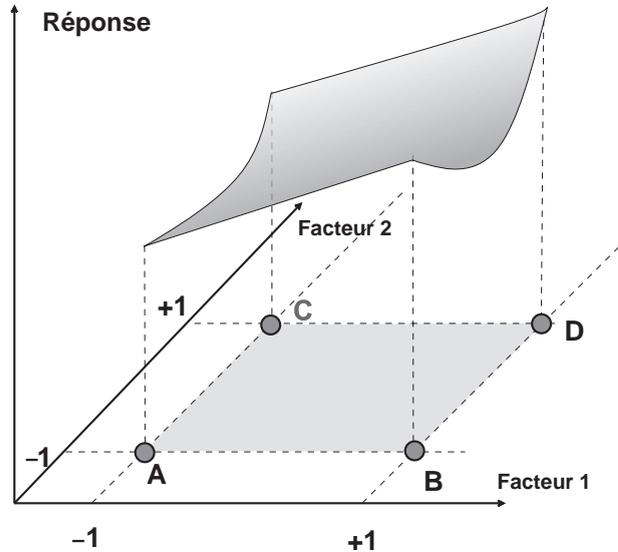


Figure 1.10 – L'ensemble des réponses qui correspond à tous les points du domaine d'étude forme la surface de réponse.

En général, on ne connaît que quelques réponses, celles qui correspondent aux points expérimentaux retenus par l'expérimentateur. On interpole à l'aide d'un modèle mathématique, les réponses inconnues pour obtenir la surface de réponse. Les points d'expériences retenus par la théorie des plans d'expériences assurent la meilleure précision possible sur la forme et la position de la surface de réponse.

1.6.4 Modélisation mathématique *a priori* de la réponse

■ Modélisation mathématique

En l'absence de toute information sur la fonction qui lie la réponse aux facteurs, on se donne *a priori* une loi d'évolution dont la formulation la plus générale est la suivante :

$$y = f(x_1, x_2, x_3, \dots, x_n) \quad \{1.2\}$$

Cette fonction est trop générale et il est d'usage d'en prendre un développement limité de Taylor-Mac Laurin, c'est-à-dire une approximation. Si les dérivées peuvent être considérées comme des constantes, le développement précédent prend la forme d'un polynôme de degré plus ou moins élevé :

$$y = a_0 + \sum a_i x_i + \sum a_{ij} x_i x_j + \sum a_{ii} x_i^2 + \dots \quad \{1.3\}$$

où :

- y est la grandeur à laquelle s'intéresse l'expérimentateur ; c'est la réponse ou la grandeur d'intérêt,
- x_i représente un niveau du facteur i ,
- x_j représente un niveau du facteur j ,
- a_0, a_i, a_{ij}, a_{ii} sont les *coefficients* du polynôme.

Ce modèle est appelé le modèle *a priori* ou le modèle *postulé*.

Les modèles établis sont des modèles de prévision valables dans le domaine d'étude, domaine que l'on doit toujours préciser. Ce ne sont pas des modèles théoriques basés sur des lois physico-chimiques ou mécaniques. Dans quelques rares cas, il est possible d'utiliser des lois physiques théoriques connues.

■ Modélisation expérimentale

Deux compléments doivent être apportés au modèle purement mathématique précédemment décrit.

Le premier complément est le manque d'ajustement. Cette expression traduit le fait que le modèle choisi par l'expérimentateur avant les expériences est probablement un peu différent du modèle réel qui régit le phénomène étudié. Il y a un écart entre ces deux modèles. Cet écart est le *manque d'ajustement* (*lack of fit* en anglais), on le note par la lettre Δ .

Le second complément est la prise en compte de la nature aléatoire de la réponse. En effet, dans le cas général, si l'on mesure plusieurs fois une réponse en un même point expérimental, on n'obtiendra pas exactement le même résultat. Il y a une dispersion des résultats. Les dispersions ainsi constatées sont appelées *erreurs aléatoires* ou *erreurs expérimentales* (*pure error* en anglais) et on les note par la lettre ε .

La relation générale {1.2} doit être modifiée ainsi :

$$y = f(x_1, x_2, x_3, \dots, x_n) + \Delta + \varepsilon \quad \{1.4\}$$

Cette relation sera exploitée au cours du chapitre 5 où l'on verra comment on estime le manque d'ajustement Δ et l'erreur aléatoire ε .

■ Système d'équations

Chaque point expérimental apporte une valeur de la réponse. Or cette réponse est modélisée par un polynôme dont les coefficients sont les inconnues qu'il faut déterminer. À la fin du plan d'expériences, on a un système de n équations (s'il y a n essais) à p inconnues (s'il y a p coefficients dans le modèle choisi *a priori*). Ce système s'écrit d'une manière simple en notation matricielle :

$$y = X a + e \quad \{1.5\}$$

où :

- y est le *vecteur des réponses*,
- X est la *matrice de calcul des coefficients* ou *matrice du modèle* qui dépend des points expérimentaux choisis pour exécuter le plan et du modèle postulé,
- a est le *vecteur des coefficients*,
- e est le *vecteur des écarts*.

Ce système ne peut pas, en général, être résolu simplement car le nombre d'équations est inférieur au nombre d'inconnues. En effet, il y a n équations et $p + n$ inconnues. Cette résolution ne peut être menée à bien que si l'on utilise une méthode de régression. La plupart du temps cette méthode est basée sur le critère d'optimisation des moindres carrés. On obtient ainsi les estimations des coefficients que l'on note

$$\hat{\mathbf{a}}$$

Le résultat de ce calcul est :

$$\hat{\mathbf{a}} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y} \quad \{1.6\}$$

formule dans laquelle la matrice \mathbf{X}' est la matrice transposée de \mathbf{X} (voir l'annexe D sur le calcul matriciel). Il existe de nombreux logiciels qui exécutent ce calcul et qui donnent directement les valeurs des coefficients.

Deux matrices interviennent constamment dans la théorie des plans d'expériences :

- la matrice d'information $\mathbf{X}' \mathbf{X}$,
- la matrice de dispersion $(\mathbf{X}' \mathbf{X})^{-1}$.

2 • VOTRE PREMIER PLAN D'EXPÉRIENCES

Si vous avez lu le premier chapitre, vous en savez assez pour attaquer votre premier plan d'expériences. Nous allons suivre la démarche suivante dont les principales étapes sont :

- la préparation de l'étude,
- le choix du plan d'expériences,
- l'expérimentation,
- l'interprétation des résultats,
- l'arrêt ou la poursuite de l'étude.

Nous allons examiner ces différentes phases sur un premier exemple.

2.1 Exemple 01 : je maîtrise la consommation de ma voiture

2.1.1 Préparation du plan d'expériences

Cette étape se décompose en plusieurs parties dont les principales sont décrites dans les paragraphes suivants.

■ Définition de l'objectif de l'étude

Nous allons suivre la réalisation d'une étude que tout le monde peut mettre sur pieds. L'objectif de l'étude est de connaître la consommation d'essence d'une voiture lorsqu'on roule plus ou moins chargé et plus ou moins vite.

L'étude sera réussie si on peut répondre à des questions telles que : « Combien ma voiture consomme-t-elle : si elle est à vide ou non ? si je roule vite ou non ? si elle est chargée et si je roule vite ? »

■ Description des éléments sur lesquels va porter l'expérimentation

Je vais décrire les essais que j'ai effectués sur ma propre voiture. Je vais être le seul conducteur et j'effectuerai les essais sur une autoroute. Si j'en avais la possibilité, je ferais les essais sur un circuit automobile où il est facile de respecter la plupart des conditions imposées à l'expérimentation. Il est en effet plus facile de réaliser des essais sur un circuit, dans un laboratoire, un atelier pilote ou tout lieu spécialement conçu pour ce genre d'activité.

■ Choix de la réponse permettant d'atteindre l'objectif

La réponse sera la consommation d'essence en litres pour 100 kilomètres parcourus.

■ Recherche des facteurs qui pourraient être influents sur la réponse

Quels sont les facteurs qui peuvent modifier la consommation d'une voiture ?

Il y a bien sûr les deux facteurs que je vais étudier, la charge et la vitesse de la voiture. Mais il y en a d'autres. Par exemple, la marque et la pression de gonflage des pneumatiques, la présence ou non d'une galerie, la direction du vent, la pluie, le réglage du moteur, le nombre d'arrêts et de démarrages, la route plus ou moins accidentée. Il est prudent de noter par écrit tous les facteurs possibles. Pour remplir la liste des facteurs et essayer de ne pas en oublier, on peut aller voir son garagiste pour lui demander s'il ne connaît pas d'autres facteurs pouvant augmenter ou réduire la consommation. On a toujours intérêt à bien faire le tour du problème. On ne sera pas obligé d'étudier tous les facteurs mais on connaîtra ceux qui pourraient être influents. Les facteurs qui ne seront pas étudiés dans le plan d'expériences seront, en général, fixés à un niveau constant pendant toute l'expérimentation.

■ Définition des niveaux des facteurs

Il s'agit de choisir les niveaux haut et bas de chaque facteur.

La vitesse ne devra pas être trop faible et elle ne devra pas dépasser les limitations imposées par la sécurité. Dans notre exemple, le niveau bas sera de 80 km/h et le niveau haut de 120 km/h.

La charge à vide est définie par la présence du conducteur seul, sans aucun bagage. La surcharge est définie par le poids ajouté à la charge à vide. S'il y a 3 personnes supplémentaires et 25 kg de bagage par personne, la surcharge est d'environ $3 \times 70 = 210$ pour les passagers et $4 \times 25 = 100$ kg pour les bagages des quatre personnes à bord, soit environ 300 kg de surcharge. Au lieu de faire appel à ses voisins pour les essais, on pourra utiliser des gueuses de fonte régulièrement réparties sur les sièges et dans le coffre. Le poids total des gueuses est de 300 kg.

Le niveau bas du facteur charge sera : la voiture et le conducteur seul et sans bagage.

Le niveau haut du facteur charge sera : la voiture, le conducteur et une charge supplémentaire de 300 kg régulièrement répartie.

Il est bon de résumer ces niveaux dans un tableau (Tableau 2.1).

Tableau 2.1 – Facteurs et domaine d'étude.

Facteur	Niveau bas (-)	Niveau haut (+)
Vitesse (1)	80 km/h	120 km/h
Surcharge (2)	0	300 kg

Ce tableau est important car les conclusions de l'expérimentation ne seront valables qu'à l'intérieur de ce domaine d'étude. Si l'on charge la voiture à 400 kg et que l'on roule à 130 km/h, on ne pourra pas utiliser les conclusions de cette étude.

■ Examen des contraintes

Lorsque l'on roule, la quantité d'essence diminue. Le poids du véhicule est donc modifié au cours de l'expérimentation. Il faudra commencer chaque essai avec le réservoir plein pour que le facteur « diminution d'essence au cours de l'essai » n'ait pas trop d'influence. Cela oblige à débiter le parcours dans une station service.

Il faudra qu'il fasse le même temps pour tous les essais (température, vent, pluie). On choisira donc de faire les essais le même jour et, si possible, un jour de beau temps sans vent ni pluie. La longueur du parcours devra être compatible avec cette contrainte.

On choisira le même parcours et dans le même sens pour chaque essai pour que l'influence des côtes et des descentes soit toujours la même.

On vérifiera que les pneus sont bien gonflés à la même pression avant chaque essai. On peut ainsi faire une liste de précautions à prendre pour tenir compte des contraintes.

2.1.2 Choix du plan d'expériences

On sait qu'il y a deux facteurs à étudier. Les niveaux bas et haut de chaque facteur ont été définis. Les facteurs à conserver constants pendant l'expérimentation sont : le parcours et par conséquent le kilométrage, le sens du parcours, le point de départ, le point d'arrivée et la pression des pneus.

Ayant deux facteurs prenant chacun deux niveaux, le plus simple est de choisir un plan d'expériences factoriel complet 2^2 . La dénomination 2^2 a la signification suivante : le 2 en exposant indique le nombre de facteurs, l'autre 2 indique les nombres de niveaux des facteurs. Ce plan est bien adapté à notre problème puisqu'il correspond exactement à deux facteurs prenant chacun deux niveaux. Les points d'expériences ont pour coordonnées les niveaux bas et les niveaux hauts des facteurs.

On peut présenter ce plan de plusieurs manières qui se complètent les unes les autres. On peut d'abord dessiner le domaine d'étude dans l'espace expérimental, puis ajouter les points d'expériences en tenant compte de leurs coordonnées (Figure 2.1).

On peut aussi représenter les expériences à faire sous forme de tableaux, en utilisant soit les grandeurs habituelles ou légales (km/h et kg), soit les grandeurs codées. Avec les grandeurs légales, le tableau prend le nom de *tableau d'expérimentation* ou de *matrice d'expérimentation*. Avec les grandeurs codées, le tableau prend le nom de *plan d'expériences* ou de *matrice d'expériences*.

La première colonne de la matrice d'expérimentation est utilisée pour indiquer les noms des essais (Tableau 2.2). On peut soit les numéroter, soit leur donner un nom. La deuxième colonne est celle du premier facteur, on y indique successivement les niveaux qu'il faut donner à ce facteur. La troisième colonne est celle du deuxième facteur et on y indique également les niveaux de ce facteur. Le premier essai, essai n° 1 ou essai A, sera exécuté avec une vitesse de 80 km/h et sans surcharge. Le deuxième essai, essai n° 2 ou essai B, sera exécuté avec une vitesse de 120 km/h et sans surcharge. Le troisième essai, essai n° 3 ou essai C, sera exécuté avec une vitesse de 80 km/h et avec une surcharge de 300 kg. Enfin le quatrième essai, essai n° 4 ou essai D, sera exécuté avec une vitesse de 120 km/h et avec une surcharge de 300 kg. Ce tableau est très utile pour l'exécution des essais.

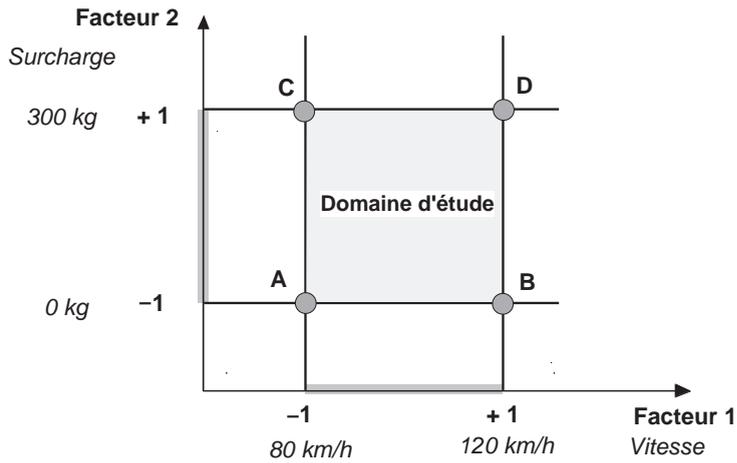


Figure 2.1 – Représentation du plan d'expérimentation.

Tableau 2.2 – Matrice d'expérimentation.

Essai n°	Vitesse Facteur 1	Surcharge Facteur 2
1 (A)	80 km/h	0 kg
2 (B)	120 km/h	0 kg
3 (C)	80 km/h	300 kg
4 (D)	120 km/h	300 kg

Tableau 2.3 – Matrice d'expériences.

Essai n°	Vitesse Facteur 1	Surcharge Facteur 2
1 (A)	-1	-1
2 (B)	+1	-1
3 (C)	-1	+1
4 (D)	+1	+1
Niveau -1	80 km/h	0 kg
Niveau +1	120 km/h	300 kg

La première colonne de la matrice d'expériences (Tableau 2.3) est utilisée de la même manière pour indiquer les noms des essais. La deuxième colonne est celle du premier facteur, on y indique successivement les niveaux qu'il faut donner à ce

facteur mais cette fois sous forme codée c'est-à-dire avec des -1 et des $+1$. La troisième colonne est celle des niveaux codés du deuxième facteur. On trouve deux lignes en bas du tableau pour indiquer la signification des niveaux -1 et $+1$ de chaque facteur. Ce tableau est utilisé lors de la construction du plan.

Pour l'interprétation des résultats d'essais, on utilise soit la matrice d'expérimentation, soit la matrice d'expériences selon les faits que l'on veut mettre en évidence.

2.1.3 Expérimentation

C'est la partie technique de l'étude. Il faut que les essais soient menés avec beaucoup de soin et il faut être sûr de la qualité des résultats. Pour le présent exemple, l'ordre des essais a peu d'importance, mais nous verrons que ce n'est pas toujours le cas. Nous étudierons au chapitre 8 les principes qui régissent le choix de l'ordre des essais. Dans le cadre de l'étude en cours, il est facile de changer la vitesse d'un essai à l'autre. En revanche, il serait maladroit de charger la voiture pour le premier essai, de la décharger pour le second et de la recharger pour le troisième ou le quatrième essai. On fera donc, en premier, les essais n° 1 et 2, puis les essais n° 3 et 4. On pourrait aussi faire d'abord les essais n° 3 et 4, puis les essais n° 1 et 2. Dans les deux cas, il n'y a qu'un seul chargement à faire.

Le parcours a été choisi et fait 120 km de long, soit 5 heures de conduite. Les temps d'arrêt à la pompe à essence, de mesure de la consommation, de la vérification de la pression des pneumatiques et de chargement de la voiture sont estimés à environ 2 heures. On pourra faire les essais en une journée.

On exécute l'essai n° 1 (80 km/h et sans surcharge) et on note la consommation d'essence. On trouve 10 litres soit une consommation de 8,3 litres aux 100 kilomètres. Le deuxième essai (120 km/h et sans surcharge) donne une consommation de 10,7 litres aux 100 kilomètres. Les deux autres essais (essais n° 3 et 4) fournissent respectivement une consommation de 9,7 et de 12,3 litres aux 100 kilomètres. Les résultats sont consignés dans une quatrième colonne de la matrice d'expérimentation ou de la matrice d'expériences (Tableau 2.4).

Tableau 2.4 – Matrice d'expériences et résultats.

N° essai	Vitesse <i>Facteur 1</i>	Surcharge <i>Facteur 2</i>	Consommation (l/100 km)
1 (A)	-1	-1	8,3
2 (B)	+1	-1	10,7
3 (C)	-1	+1	9,7
4 (D)	+1	+1	12,3
Niveau -1	80 km/h	0 kg	
Niveau +1	120 km/h	300 kg	

On peut reporter ces résultats sur le domaine d'étude (Figure 2.2).

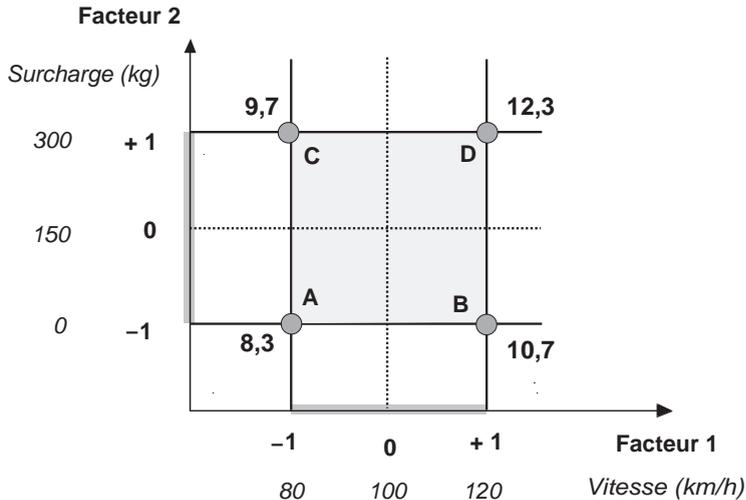


Figure 2.2 – Valeur de la réponse en divers points du domaine d'étude.

2.1.4 Signification des coefficients

Ces consommations étant mesurées, il ne semble pas facile d'en donner une interprétation simple. En effet on aimerait répondre à des questions comme : « Quelle est la consommation de ma voiture quand je la charge à 100 kg et que je roule à 90 km/h ? » ou « Quelle est la consommation supplémentaire quand je roule à 100 km/h au lieu de 80 km/h avec une surcharge de 150 kg ? » Les réponses à ces questions existent. Elles se cachent dans les résultats que nous avons enregistrés, mais pour le moment nous ne les voyons pas. Il va falloir que nous transformions nos résultats bruts pour en avoir une vision claire et précise. C'est la phase d'interprétation. Les logiciels de plans d'expériences vont maintenant prendre de l'importance car ce sont eux qui vont nous aider à réaliser toutes les transformations qui nous sont nécessaires et toutes les illustrations qui nous aideront à bien comprendre les résultats.

Le modèle postulé des plans factoriels complets 2^2 est :

$$y = a_0 + a_1x_1 + a_2x_2 + a_{12}x_1x_2 \quad \{2.1\}$$

où :

- y est la réponse, dans cet exemple, la consommation d'essence ;
- x_1 représente le niveau du facteur 1 (la vitesse), dans cet exemple 80 km/h (ou -1) et 120 km/h (ou +1) selon les essais ;
- x_2 représente le niveau du facteur 2 (la surcharge), dans cet exemple 0 kg (ou -1) et 300 kg (ou +1) selon les essais ;
- x_1x_2 est le produit des niveaux des facteurs 1 et 2 ; dans cet exemple, en unités codées, ce produit est égal à -1 ($x_1x_2 = -1 \times +1 = +1 \times -1 = -1$) ou à +1 ($x_1x_2 = -1 \times -1 = +1 \times +1 = +1$) ;

- a_0 est le coefficient constant du modèle ;
- a_1 est le coefficient du facteur 1 ;
- a_2 est le coefficient du facteur 2 ;
- a_{12} est le coefficient du terme x_1x_2 .

Ce modèle est appelé modèle *polynomial du premier degré avec interactions* ou modèle *PDAI* et nous allons examiner la signification de ses coefficients.

■ **Signification du coefficient constant**

Pour trouver la signification du coefficient constant a_0 , il suffit de donner la valeur 0 (unités codées) aux niveaux des deux facteurs. Le point représentatif de l'expérience correspondante est alors au centre du domaine d'étude (Figure 2.3) et la réponse en ce point a pour valeur y_0 .

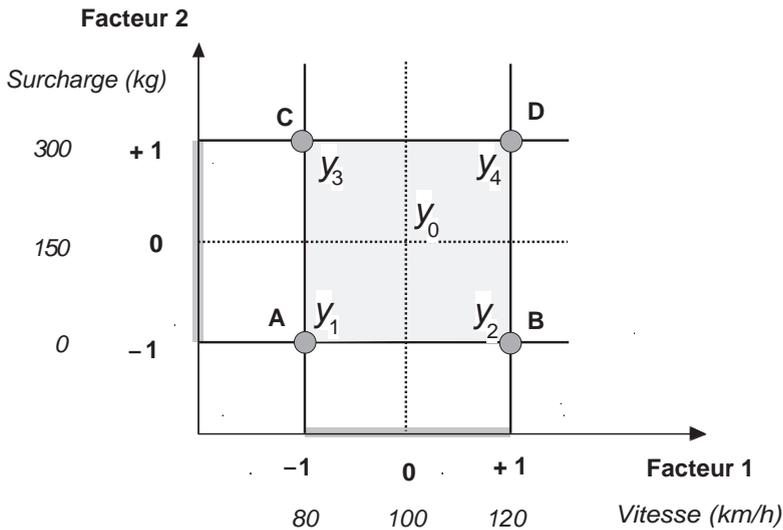


Figure 2.3 – Le coefficient constant a pour valeur la réponse au centre du domaine d'étude.

La relation {2.1} devient :

$$y = a_0 + a_1 \times 0 + a_2 \times 0 + a_{12} \times 0 \times 0$$

$$y_0 = a_0$$

La valeur du coefficient constant a_0 est égale à la réponse au centre du domaine d'étude.

■ **Signification du coefficient du facteur 1**

Considérons les deux points B et D qui se trouvent au niveau haut du facteur 1. Les coordonnées de ces points sont, en unités codées :

B	$x_1 = +1$		D	$x_1 = +1$
	$x_2 = -1$			$x_2 = +1$

La réponse au point B est y_2 , réponse que l'on peut écrire en remplaçant les niveaux par leurs valeurs en unités codées :

$$y_2 = a_0 + a_1 \times (+1) + a_2 \times (-1) + a_{12} \times (+1) \times (-1) = a_0 + a_1 - a_2 - a_{12}$$

La réponse au point D est y_4 , que l'on peut écrire en remplaçant les niveaux par leurs valeurs en unités codées :

$$y_4 = a_0 + a_1 \times (+1) + a_2 \times (+1) + a_{12} \times (+1) \times (+1) = a_0 + a_1 + a_2 + a_{12}$$

Additionnons les deux réponses y_2 et y_4 :

$$y_2 + y_4 = 2(a_0 + a_1)$$

Faisons le même calcul pour les points A et C qui se trouvent au niveau bas du facteur 1 et où les réponses sont respectivement y_1 et y_3 . On obtient :

$$y_1 + y_3 = 2(a_0 - a_1)$$

Si on soustrait ces deux dernières relations, on a :

$$4a_1 = -y_1 + y_2 - y_3 + y_4$$

relation que l'on peut écrire :

$$a_1 = \frac{1}{2} \left[\frac{y_2 + y_4}{2} - \frac{y_1 + y_3}{2} \right]$$

Or $\frac{y_2 + y_4}{2}$ est la moyenne des réponses au niveau haut du facteur 1. On nomme cette moyenne \bar{y}_+ . Quant à l'expression $\frac{y_1 + y_3}{2}$, c'est la moyenne des réponses au niveau bas du facteur 1, soit \bar{y}_- . On peut écrire :

$$a_1 = \frac{1}{2} [\bar{y}_+ - \bar{y}_-]$$

Le coefficient a_1 est donc la demi-différence entre la moyenne des réponses au niveau haut du facteur 1 et la moyenne des réponses au niveau bas du même facteur 1.

Quand on passe du niveau bas au niveau haut, la réponse varie, en moyenne, comme la différence $[\bar{y}_+ - \bar{y}_-]$. Si cette différence est grande, la réponse varie beaucoup, si cette différence est faible, la réponse varie peu. On a donc là un moyen de savoir comment la réponse varie en fonction du facteur 1. C'est la raison pour laquelle on appelle le coefficient a_1 l'*effet* du facteur 1.

Application

On connaît les quatre réponses, on peut donc calculer facilement le coefficient :

$$a_1 = \frac{1}{4} [-y_1 + y_2 - y_3 + y_4]$$

$$a_1 = \frac{1}{4} [-8,3 + 10,7 - 9,7 + 12,3] = 1,25$$

L'effet de la vitesse (facteur 1) est de 1,25 litre aux 100 kilomètres. Cela signifie que, si la vitesse passe de 80 à 100 km/h, la consommation augmente, en moyenne, de 1,25 litre aux 100 kilomètres. Si la vitesse passe de 80 à 120 km/h, la consommation augmente de 2,50 litres aux 100 kilomètres.

■ Représentation du coefficient du facteur 1

La moyenne des réponses au niveau haut du facteur 1, \bar{y}_+ , est située sur la surface de réponse et se trouve à l'aplomb du point M_+ , milieu du segment BD (Figure 2.4). Il a donc pour coordonnées :

$$M_+ \left| \begin{array}{l} x_1 = +1 \\ x_2 = 0 \end{array} \right.$$

La moyenne des réponses au niveau bas du facteur 1 est située sur la surface de réponse et se trouve à l'aplomb du point M_- , milieu du segment AC. Il a donc pour coordonnées :

$$M_- \left| \begin{array}{l} x_1 = -1 \\ x_2 = 0 \end{array} \right.$$

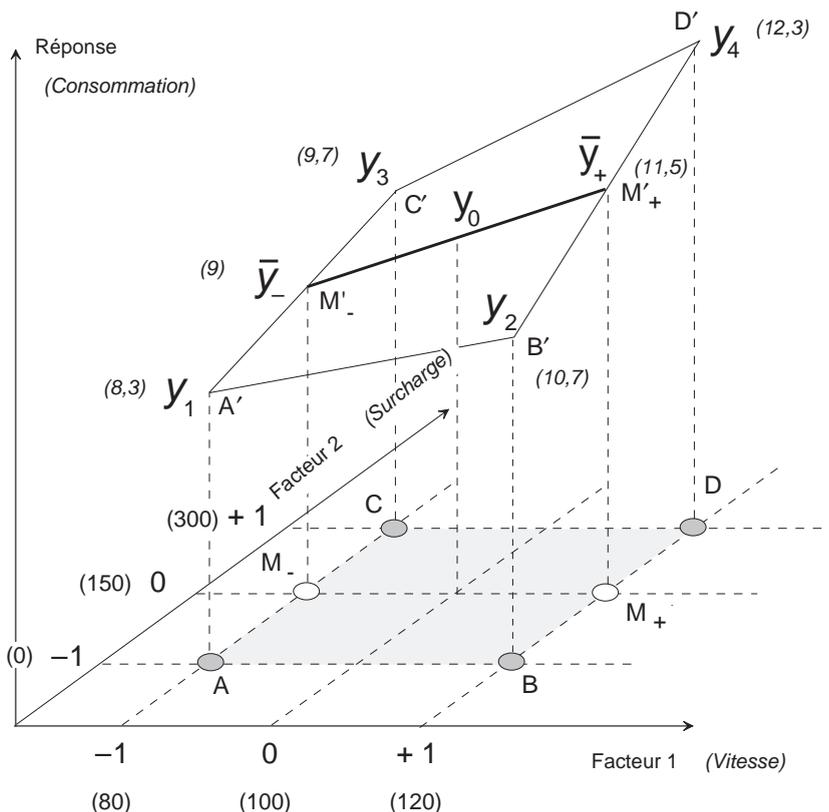


Figure 2.4 – Le coefficient du facteur 1 est la pente de la droite qui joint les deux réponses \bar{y}_+ et \bar{y}_- .

La variation de la réponse entre la moyenne des réponses au niveau haut du facteur 1, \bar{y}_+ , et la moyenne des réponses au niveau bas de ce même facteur \bar{y}_- est $\bar{y}_+ - \bar{y}_-$, c'est-à-dire deux fois le coefficient a_1 .

Le coefficient a_1 est donc égal à la variation de la réponse entre y_0 , réponse au centre du domaine d'étude, et y_+ , moyenne des réponses au niveau haut du facteur 1. On peut également regarder le coefficient a_1 comme la pente de la droite $M'_- M'_+$. On peut dire aussi que le coefficient a_1 est égal à la variation moyenne de la réponse quand le facteur 1 passe du niveau zéro au niveau haut. Il représente donc l'influence du facteur 1 dans le domaine d'étude.

■ Illustration de l'effet du facteur 1

On a l'habitude pour faire apparaître clairement la droite $M'_- M'_+$ et illustrer l'effet du facteur 1, d'extraire le plan $M_+ M_- M'_- M'_+$ de la figure 2.4. On obtient la figure 2.5 qui est beaucoup plus facile à lire.

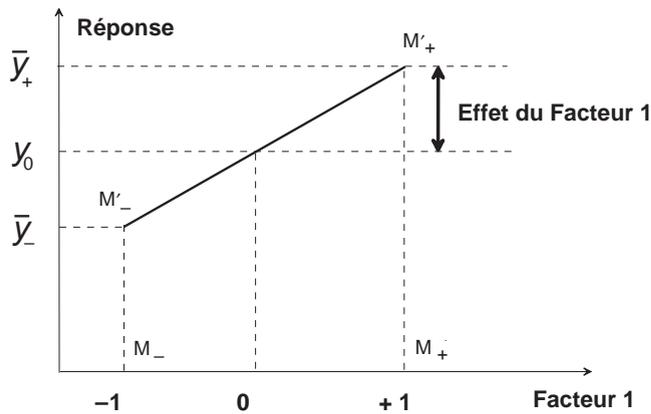


Figure 2.5 – Illustration de l'effet du facteur 1.

L'effet du facteur 1 peut être positif ou négatif suivant le signe du coefficient.

Application

Consommation (l / 100 km)

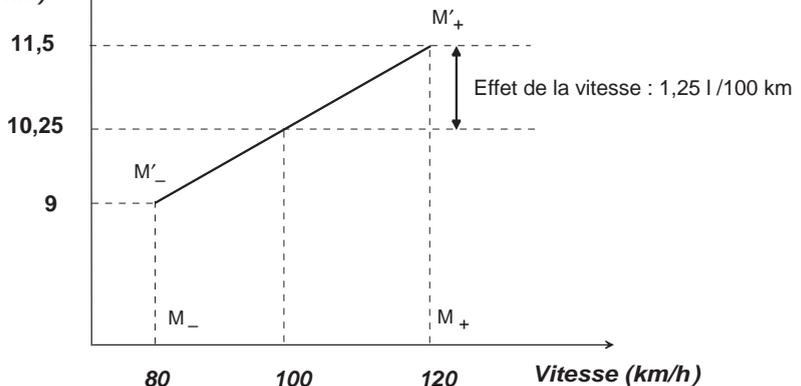


Figure 2.6 – Illustration de l'effet de la vitesse.

La moyenne des consommations au niveau haut de la vitesse (facteur 1) est :

$$\bar{y}_+ = \frac{y_2 + y_4}{2} = \frac{10,7 + 12,3}{2} = 11,5$$

La moyenne des consommations au niveau bas de la vitesse (facteur 1) est :

$$\bar{y}_- = \frac{y_1 + y_3}{2} = \frac{8,3 + 9,7}{2} = 9$$

La consommation passe, en moyenne, de 9 à 11,5 litres aux 100 kilomètres quand la vitesse passe de 80 à 120 km/h. La consommation au centre est la moitié des deux moyennes, soit 10,25 litres aux 100 kilomètres.

La consommation passe, en moyenne, de 10,25 à 11,5 litres aux 100 kilomètres quand la vitesse passe de 100 à 120 km/h. Cette augmentation de 1,25 litre aux 100 kilomètres représente l'effet du facteur « vitesse » (Figure 2.6).

■ Signification du coefficient du facteur 2

De la même manière, on montre que le coefficient a_2 est égal à la variation moyenne de la réponse quand le facteur 2 passe du niveau zéro au niveau haut. Il représente l'influence du facteur 2 dans le domaine d'étude. On l'appelle « effet du facteur 2 ».

D'une manière générale, quand le modèle choisi est un polynôme, les coefficients des termes du premier degré sont les effets des facteurs.

Application

On connaît les quatre réponses, on peut donc calculer facilement le coefficient a_2 :

$$a_2 = \frac{1}{4}[-y_1 - y_2 + y_3 + y_4]$$

$$a_2 = \frac{1}{4}[-8,3 - 10,7 + 9,7 + 12,3] = 0,75$$

L'effet de la surcharge (facteur 2) est de 0,75 l aux 100 km. Cela signifie que, si la surcharge passe de 0 à 150 kg, la consommation augmente en moyenne de 0,75 l aux 100 km. Si la surcharge passe de 0 à 300 kg la consommation augmente de 1,50 l aux 100 km.

Consommation (l / 100 km)

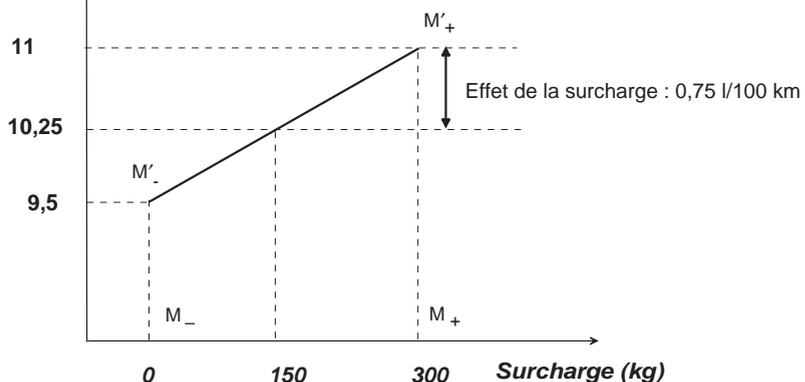


Figure 2.7 – Illustration de l'effet de la surcharge.

La moyenne des consommations au niveau haut de la surcharge (facteur 2) est :

$$\bar{y}_+ = \frac{y_3 + y_4}{2} = \frac{9,7 + 12,3}{2} = 11$$

La moyenne des consommations au niveau bas de la surcharge (facteur 2) est :

$$\bar{y}_- = \frac{y_1 + y_2}{2} = \frac{8,3 + 10,7}{2} = 9,5$$

La consommation passe, en moyenne, de 9,5 à 11 l aux 100 km quand la surcharge passe de 0 à 300 kg. La consommation au centre est la moitié des deux moyennes, soit 10,25 l aux 100 km.

La consommation passe, en moyenne, de 10,25 à 11 l aux 100 km quand la surcharge passe de 150 à 300 kg. Cette augmentation de 0,75 l aux 100 km est l'effet du facteur surcharge (Figure 2.7).

■ Signification du coefficient a_{12}

On peut calculer le coefficient a_{12} par une méthode analogue à celle qui a été utilisée pour les coefficients a_1 et a_2 . On trouve que le coefficient a_{12} est égal à :

$$a_{12} = \frac{1}{2} \left[\frac{y_4 - y_3}{2} - \frac{y_2 - y_1}{2} \right]$$

Or $\frac{y_4 - y_3}{2}$ est l'effet du facteur 1 lorsque le facteur 2 est au niveau haut. C'est la moitié de la variation de la réponse entre y_4 et y_3 . Cet effet est illustré par la pente de la droite C'D' (Figure 2.4 et Figure 2.8).

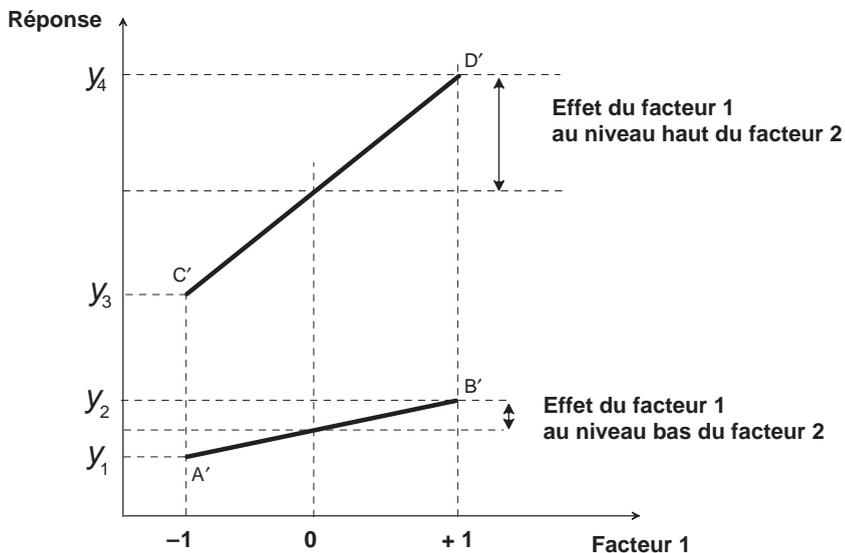


Figure 2.8 – Illustration d'une interaction entre deux facteurs.

L'expression $\frac{y_2 - y_1}{2}$ est l'effet du facteur 1 lorsque le facteur 2 est au niveau bas. C'est la moitié de la variation de la réponse entre y_2 et y_1 . Cet effet est illustré par la pente de la droite A'B' (Figure 2.8).

Le coefficient a_{12} est la moitié de la différence entre ces deux effets.

Le coefficient a_{12} mesure donc la variation de l'effet du facteur 1 quand le niveau du facteur 2 est modifié. On peut aussi montrer que le même coefficient a_{12} mesure également la variation de l'effet du facteur 2 quand le niveau du facteur 1 est, lui aussi, modifié.

Le coefficient a_{12} est appelé l'*interaction* entre les facteurs 1 et 2.

On peut illustrer une interaction entre deux facteurs en extrayant de la figure 2.4 les plans ABA'B' (niveau bas du facteur 2), et CDC'D' (niveau haut du facteur 2) et en projetant ces plans sur un même plan (Figure 2.8).

S'il n'y a pas d'interaction entre deux facteurs, les pentes des droites A'B' et C'D' sont les mêmes.

S'il y a interaction entre deux facteurs, les pentes des deux droites précédentes ne sont pas les mêmes. L'interaction est d'autant plus forte que les pentes sont différentes.

Application

On peut donc calculer l'interaction avec la relation :

$$a_{12} = \frac{1}{4}[+y_1 - y_2 - y_3 + y_4]$$

$$a_{12} = \frac{1}{4}[+8,3 - 10,7 - 9,7 + 12,3] = 0,05$$

L'interaction entre les facteurs 1 et 2 est de 0,05 l aux 100 km. Cela signifie que l'effet de la vitesse est un peu plus élevé quand on se trouve en surcharge. Quand la vitesse est de 80 km/h, l'effet de la surcharge est de 0,7 l aux 100 km. Quand la vitesse est de 120 km/h, l'effet de la surcharge est de 0,8 l aux 100 km.

Cela signifie aussi que l'effet de la surcharge est plus important quand on roule vite. Quand il n'y a pas de surcharge, l'effet de la vitesse est de 1,2 l aux 100 km. Quand la surcharge est de 300 kg, l'effet de la vitesse est de 1,3 l aux 100 km (Figure 2.9).

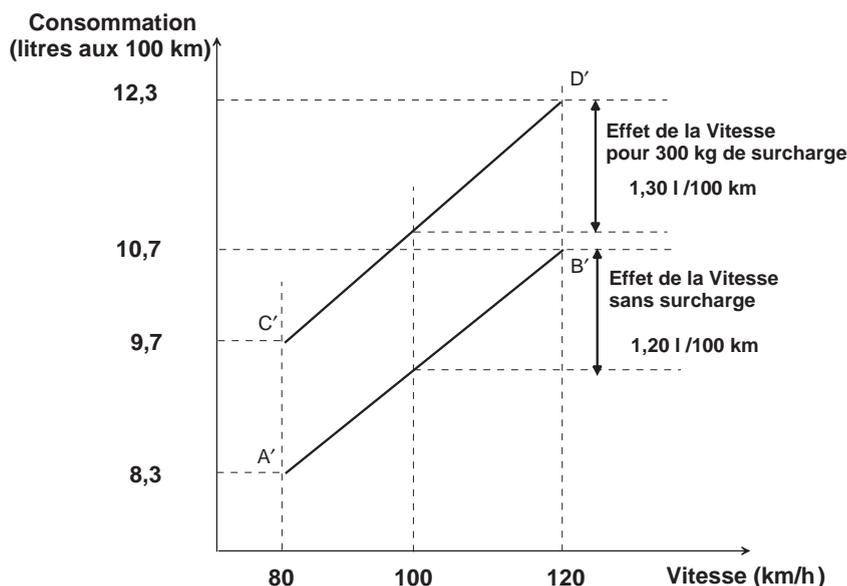


Figure 2.9 – Illustration de l'interaction entre la vitesse et la surcharge. Cette interaction est très faible puisque les deux droites ont presque la même pente.

2.1.5 Interprétation des résultats des calculs

On connaît maintenant, les valeurs :

- du coefficient constant : $a_0 = 10,25$;
- du coefficient du facteur 1 (vitesse) : $a_1 = 1,25$;
- du coefficient du facteur 2 (surcharge) : $a_2 = 0,75$;
- de l'interaction a_{12} entre la vitesse et la surcharge : $a_{12} = 0,05$.

On peut reporter ces valeurs dans la relation {2.1} du modèle postulé :

$$y = 10,25 + 1,25 x_1 + 0,75 x_2 + 0,05 x_1 x_2 \quad \{2.2\}$$

On comprend qu'avec ce modèle, on puisse calculer toutes les réponses dans le domaine d'étude. Il suffit d'attribuer des valeurs aux niveaux x_1 et x_2 pour obtenir immédiatement la consommation. Le modèle étant en unités centrées réduites, il faut faire les calculs dans ces unités et transformer ensuite les résultats obtenus en unités légales. Si l'on veut utiliser les unités légales directement, il faut transformer la relation {2.2}. Dans ce cas, il suffit d'appliquer la relation {1.1} :

$$x = \frac{A - A_0}{Pas}$$

Facteur 1 (vitesse) :

$$x_1 = \frac{A_1 - A_0}{Pas_1} = \frac{A_1 - 100}{20}$$

Facteur 2 (surcharge) :

$$x_2 = \frac{A_2 - A_0}{Pas_2} = \frac{A_2 - 150}{150}$$

$$y = 10,25 + 1,25 \left(\frac{A_1}{20} - \frac{100}{20} \right) + 0,75 \left(\frac{A_2}{150} - \frac{150}{150} \right) + 0,05 \left(\frac{A_1}{20} - \frac{100}{20} \right) \left(\frac{A_2}{150} - \frac{150}{150} \right)$$

$$y = 10,25 - \frac{125}{20} - 0,75 + \frac{1,25 A_1}{20} + \frac{0,75 A_2}{150} + \left(\frac{0,05 A_1}{20} - \frac{5}{20} \right) \left(\frac{A_2}{150} - 1 \right)$$

$$y = 3,25 + \frac{5}{20} + \left[\frac{1,25}{20} - \frac{0,05}{20} \right] A_1 + \left[\frac{0,75}{150} - \frac{5}{20 \times 150} \right] A_2 + \frac{0,05 A_1}{20} \frac{A_2}{150}$$

D'où la relation en unités légales :

$$y = 3,5 + 0,06 A_1 + 0,0033 A_2 + 0,0000167 A_1 A_2 \quad \{2.3\}$$

On peut donc répondre facilement aux questions du début du paragraphe 5 en utilisant soit la relation {2.2} soit la relation {2.3}.

Application

La question était : « Quelle est la consommation de ma voiture quand je la charge à 100 kg et que je roule à 90 km/h ? » Appliquons la relation {2.3} :

$$y = 3,5 + 0,06 \times 90 + 0,00333 \times 100 + 0,0000167 \times 90 \times 100$$

$$y = 3,5 + 5,4 + 0,333 + 0,150 = 9,383$$

La réponse est : Si je roule à 90 km/h avec une surcharge de 100 kg, je consommerai 9,38 l aux 100 km.

Application

La question était : « Quelle est la consommation supplémentaire quand je roule à 100 km/h au lieu de 80 km/h avec une surcharge de 150 kg ? » Commençons par transformer les unités légales en unités codées :

80 km/h = niveau -1

100 km/h = niveau 0

150 kg = niveau 0

et appliquons la relation {2.2} :

$$y = 10,25 + 1,25 \times (-1) + 0,75 \times (0) + 0,05(-1 \times 0)$$

$$y = 10,25 - 1,25 = 9$$

Quand je roule au niveau -1 (80 km/h) avec une surcharge au niveau 0 (150 kg), je consomme 9 litres de carburant aux 100 kilomètres.

$$y = 10,25 + 1,25 \times (0) + 0,75 \times (0) + 0,05(0 \times 0)$$

$$y = 10,25$$

Quand je roule au niveau 0 (100 km/h) avec une surcharge au niveau 0 (150 kg), je consomme 10,25 litres de carburant aux 100 kilomètres.

La réponse est donc : avec une surcharge de 150 kg, la consommation augmente de 1,25 l aux 100 km quand je passe de 80 à 100 km/h.

Il est donc possible de répondre aux questions posées ainsi qu'à beaucoup d'autres faisant intervenir la vitesse et la charge. Mais ne trouvez-vous pas que ces calculs sont fastidieux, qu'ils ne sont guère enrichissants et que notre esprit pourrait être utilisé à des tâches plus productives ? C'est pourquoi nous vous conseillons d'utiliser un logiciel de plan d'expériences. Tous les calculs que nous avons vus, et beaucoup d'autres, sont réalisés avec facilité et rapidité. Les logiciels permettent aussi de tracer des graphiques faisant ressortir les principales caractéristiques de l'étude. Ils multiplient la puissance de raisonnement de l'expérimentateur et réalisent des calculs impossibles à faire à la main. En effet, dès que le nombre de facteurs augmente, les calculs se compliquent et l'utilisation d'un logiciel facilite considérablement la tâche de l'expérimentateur.

À titre d'exemple les deux diagrammes suivants montrent la puissance des logiciels. Le premier diagramme (Figure 2.10) indique l'effet des deux facteurs, vitesse et surcharge, ainsi que la valeur correspondante de la réponse. Ce diagramme est interactif, on peut choisir les niveaux que l'on veut dans tout le domaine d'étude. Pour chaque couple de niveaux, on obtient la valeur de la réponse. On peut donc facilement répondre à la première question sans être obligé de faire le détail des calculs. Ce même diagramme permet aussi de répondre à la deuxième question.

Le deuxième diagramme (Figure 2.11) indique les courbes isoréponses dans le domaine d'étude. Par exemple, la courbe 10 indique toutes les combinaisons de vitesse et de surcharge qui engendrent une consommation de 10 l de carburant aux 100 km. Si l'on ne veut pas dépasser une consommation de 11 l aux 100 km, on voit comment il faut adapter sa vitesse en fonction de la surcharge. Par exemple si la surcharge est de 200 kg, il ne faudra pas dépasser 107,9 km/h.

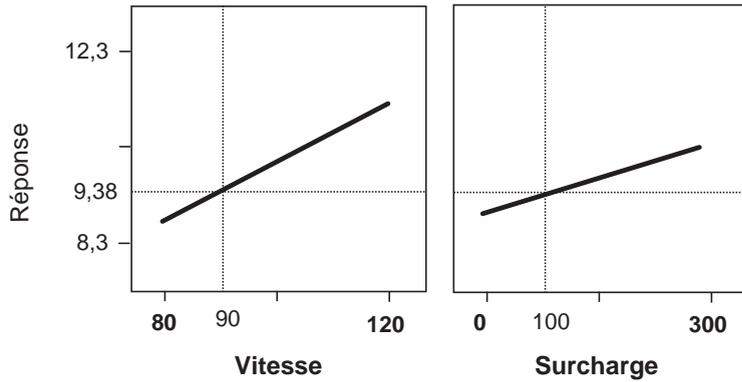


Figure 2.10 – Valeur de la réponse en divers points du domaine d'étude.

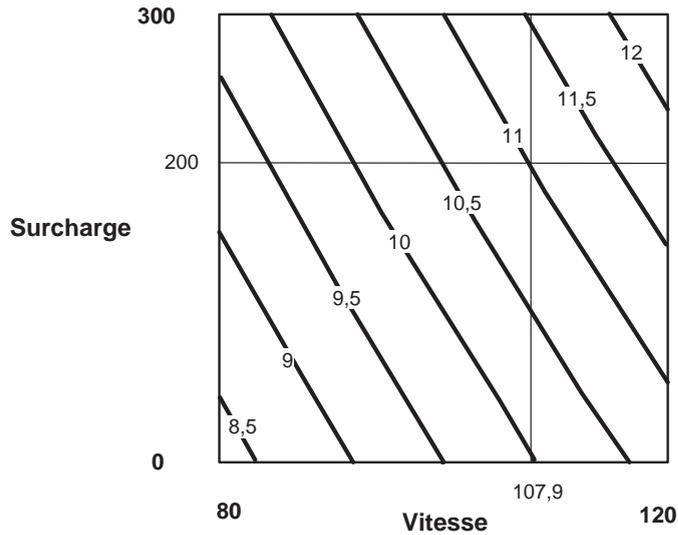


Figure 2.11 – Avec une surcharge de 200 kg on limitera la vitesse à 107,9 km/h pour ne pas consommer plus de 11 l aux 100 km.

Si vous voulez réaliser vous-même ces calculs et ces graphiques, consultez le chapitre 14 où vous trouverez les explications détaillées pour mener à bien ces opérations. Vous pouvez aussi consulter le site <http://www.plansdexpériences.com> consacré aux exemples de ce livre.

3 • PLAN D'EXPÉRIENCES POUR 3 FACTEURS

Le chapitre précédent était destiné à vous montrer que, même sur un exemple simple ne comprenant que deux facteurs, on pouvait tirer beaucoup de renseignements à partir de quatre essais seulement. Si beaucoup d'informations peuvent être extraites d'une étude faisant intervenir deux facteurs, on comprend qu'une étude de plusieurs facteurs soit encore plus fructueuse. C'est la raison pour laquelle les plans multifactoriels vont être décrits maintenant. Nous commencerons par trois facteurs et nous verrons une étude de sept facteurs. Les principes étant toujours les mêmes quel que soit le nombre de facteurs, il est tout à fait possible d'aller plus loin et on trouve dans les articles publiés ou dans les archives des sociétés des plans de plus de dix facteurs. Au-delà, les difficultés rencontrées ne concernent pas la théorie mais la mise en œuvre et le contrôle des niveaux des facteurs pendant l'expérimentation. En effet, pour onze facteurs il faut contrôler tous les niveaux à chaque essai et être absolument sûr qu'ils sont tous parfaitement respectés. Si l'on réalise 12 essais, cela fait 132 niveaux à garantir. La moindre erreur risque de compromettre l'étude. Dans le cas des simulations numériques, il est facile de programmer le passage des calculs et d'être sûr du respect des niveaux. Il n'y a alors plus aucune limite au nombre de facteurs étudiés.

3.1 Exemple 02 : des bijoux en or

3.1.1 Préparation de l'étude

■ Description de l'étude

Le grand intérêt de l'or est qu'il ne s'oxyde pas et garde toujours son brillant et sa belle couleur dorée. Mais, c'est un métal cher et c'est pourquoi beaucoup d'objets sont simplement recouverts d'une fine couche d'or qui les protège et leur donne l'apparence de l'or massif. Les bijoux plaqués or sont recouverts d'un fin dépôt d'or qui doit avoir le même aspect que l'or mais aussi des propriétés de résistance mécanique qui assurent leur longévité. Ces dépôts sont effectués par électrolyse. Le métal à recouvrir d'or est plongé dans une solution contenant un sel d'or et il est relié à la cathode d'une pile électrique (Figure 3.1). Ces dépôts doivent posséder de bonnes propriétés mécaniques pour résister aux diverses sollicitations auxquelles ils seront soumis. Pour améliorer leur qualité, il est coutume de former des alliages en ajoutant d'autres métaux en présence d'additifs. Les alliages les plus utilisés sont généralement à base de nickel ou de cobalt.

Dans la présente étude, on examine les alliages d'or et de cobalt. Pour cela on réalise une solution d'électrolytes contenant des cyanures de ces deux métaux.

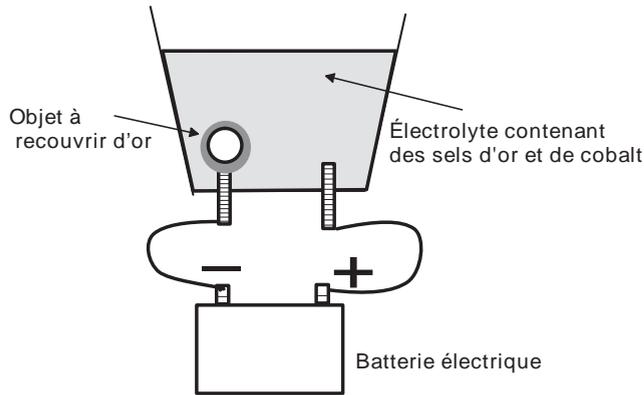


Figure 3.1 – Schéma de principe des dépôts électrolytiques d'or.

■ Définition de l'objectif de l'étude

Les objectifs de l'étude sont :

- Réaliser le dépôt le plus vite possible en respectant les impératifs de qualité.
- Obtenir une teneur en cobalt comprise entre 3 500 et 4 500 ppm dans l'alliage du dépôt.

L'étude sera réussie si on peut donner des consignes de réglage des facteurs pour atteindre les deux objectifs.

■ Choix de la réponse permettant d'atteindre l'objectif

- La première réponse correspond au premier objectif. On choisira donc la vitesse de dépôt et elle sera mesurée en milligrammes par minute (mg/min). Pour assurer la rentabilité du procédé, il faut une vitesse de dépôt au moins égale à 80 mg/min.
- La seconde réponse correspond au second objectif. On choisira donc la teneur en Cobalt de l'alliage. Cette teneur sera mesurée en partie par million (ppm). Pour assurer la qualité du dépôt, la teneur en cobalt de l'alliage doit être comprise entre 3 500 et 4 500 ppm.

■ Recherche des facteurs qui pourraient être influents sur la réponse

Quels sont les facteurs qui peuvent modifier la vitesse de dépôt et la teneur en cobalt de l'alliage ? Pour essayer de ne pas oublier de facteurs, il faut procéder méthodiquement. Il faut analyser tous les grands secteurs pouvant faire intervenir des facteurs modifiant les réponses choisies. On utilise souvent la méthode de l'arête de poisson d'Ishikawa (Figure 3.2) associée à la méthode des 6M (Méthode, Matières, Milieu, Main d'œuvre, Machines et Mesures). Cette méthode consiste à segmenter la recherche des facteurs ayant une possible influence sur les réponses. Par exemple, on se demande d'abord si les Matières premières ont une influence.

Dans le cas de la présente étude l'origine et la pureté du sel d'or peuvent modifier les résultats, la nature du sel d'or également. Il en est de même du cobalt, des additifs, de l'eau et même du courant électrique. Après l'examen des matières premières, on s'intéresse aux méthodes employées, c'est-à-dire aux conditions opératoires : intensité et voltage au cours de l'électrolyse, électrolyse en continue ou par séquences, etc. Cette quête systématique des facteurs permet de faire le tour du problème et d'éviter les gros oublis.

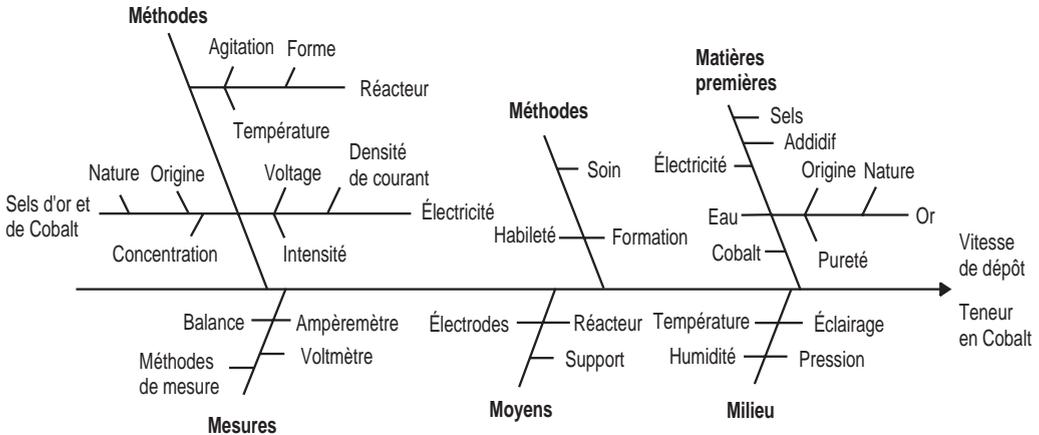


Figure 3.2 – Exemple simplifié de diagramme d'Ishikawa.

À titre d'exemple nous avons construit un diagramme d'Ishikawa réunissant quelques facteurs pouvant être influents et nous en avons donné un début de liste pour illustrer la démarche (les chercheurs qui ont réalisé cette étude en avaient fait, eux, une liste bien plus complète) :

- température de la solution d'électrolyte,
- pH de la solution d'électrolyte,
- concentration en l'or de l'électrolyte,
- nature et concentration des additifs ajoutés à l'électrolyte,
- concentration en cobalt de l'électrolyte,
- nature du support,
- surface du support,
- mouvement du support,
- position des électrodes,
- forme du réacteur,
- agitation de la solution,
- forme des électrodes,
- densité de courant, intensité, voltage, etc.

À partir de cette liste, on sélectionne les facteurs que l'on pense être influents et on les divise en deux catégories : les facteurs qui seront étudiés grâce au plan d'expériences et les facteurs qui ne seront pas étudiés au cours de l'expérimentation mais

dont les niveaux seront fixés toujours à la même valeur. Dans cette étude, trois facteurs ont été retenus pour construire le plan d'expériences :

- concentration en or de la solution d'électrolyte,
- densité de courant,
- concentration en cobalt de la solution d'électrolyte.

Les autres facteurs sélectionnés qui peuvent modifier les réponses seront fixés à un niveau constant pour tous les essais. Par exemple, on a fixé les facteurs suivants :

- le pH est fixé à 4,5 ;
- la nature et l'origine des sels ;
- la nature et l'origine des additifs ;
- le réacteur et les électrodes resteront les mêmes ;
- les conditions électriques de l'électrolyse ;
- la température du bain d'électrolyse, etc.

Les facteurs qui resteront fixes pendant l'expérimentation et les facteurs du plan d'expériences sont les *facteurs contrôlés*. Tous les autres facteurs qui ne sont pas fixés sont les *facteurs non contrôlés*. Si les facteurs non contrôlés ont peu d'influence ou s'ils restent au même niveau pendant tous les essais, ils n'entraîneront que de petites variations. Les facteurs non contrôlés sont à l'origine de l'erreur expérimentale. Si, par un mauvais hasard, un facteur influent n'a pas été pris en compte et qu'il change de niveau pendant l'expérimentation, il grossirait considérablement l'erreur expérimentale et risquerait de compromettre l'étude. Il faut donc être vigilant quand on établit la liste des facteurs pour essayer de n'en oublier aucun d'important. L'avantage de la méthode des plans d'expériences par rapport à la méthode classique, est d'obliger l'expérimentateur à réfléchir à ce problème et de lui éviter d'éventuels oublis.

■ Définition des niveaux des facteurs

Il s'agit de choisir les niveaux hauts et bas des trois facteurs retenus :

- concentration en or dans la solution d'électrolyte (en grammes par litre),
- densité de courant pendant l'électrolyse (en ampères par décimètre carré),
- concentration en cobalt dans la solution d'électrolyte (en grammes par litre).

Il est bon de résumer ces niveaux dans un tableau (Tableau 3.1).

Tableau 3.1 – Facteurs et domaine d'étude.

Facteur	Niveau bas (-)	Niveau haut (+)
Concentration en or (1)	2 g/l	15 g/l
Densité de courant (2)	5 A/dm ²	25 A/dm ²
Concentration en cobalt (3)	0,5 g/l	1,5 g/l

Nous insistons sur l'importance de bien préciser les bornes du domaine d'étude. En effet, les conclusions de l'étude ne sont valides qu'à l'intérieur de ce domaine. En dehors, on peut essayer de faire des projections de résultats mais il ne faut, en aucun cas, les considérer comme vraies et en tirer des conclusions.

3.1.2 Choix du plan d'expériences

Il y a trois facteurs à étudier. Les niveaux bas et hauts de chaque facteur ont été définis. Les niveaux des facteurs à conserver constants pendant l'expérimentation ont également été précisés et seront vérifiés avant chaque essai.

Ayant trois facteurs prenant chacun deux niveaux, et comme on pense que le modèle du premier degré avec interactions (modèle PDAI) est suffisant pour expliquer les résultats, il convient de choisir un plan d'expériences factoriel complet 2^3 :

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{23}x_2x_3 + a_{123}x_1x_2x_3 \quad \{3.1\}$$

Nous verrons que si l'on soupçonne un modèle du second degré, on choisira un plan différent. La dénomination 2^3 a la signification suivante : le 3 en exposant indique le nombre de facteurs, le 2 indique le nombre de niveaux des facteurs.

On peut représenter ce plan par une figure (Figure 3.3) en indiquant le domaine d'étude et les points d'expériences. Les points d'expériences ont pour coordonnées les niveaux bas et les niveaux hauts des facteurs. On peut également représenter ce plan par des matrices, la matrice d'expérimentation (unités normales) ou la matrice d'expériences (unités codées).

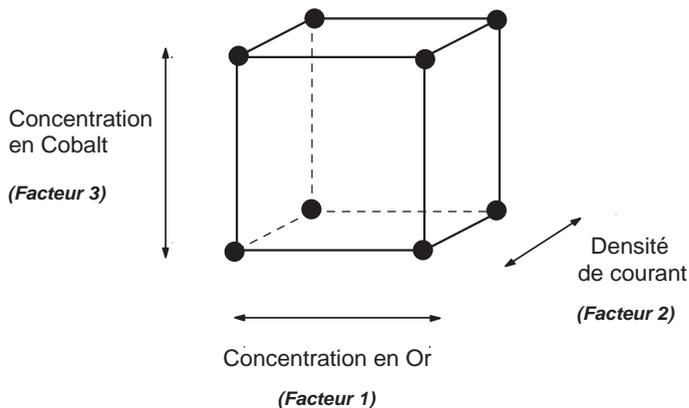


Figure 3.3 – Représentation du plan d'expériences.

La matrice d'expérimentation (Tableau 3.2) comporte 6 colonnes : les noms des essais, les trois facteurs et les deux réponses.

La matrice d'expériences (Tableau 3.3) comporte également les mêmes six colonnes, mais deux lignes ont été ajoutées pour préciser la signification des niveaux -1 et $+1$.

Pour l'interprétation des résultats d'essais, on utilise soit la matrice d'expérimentation, soit la matrice d'expériences selon les particularités que l'on veut mettre en évidence.

Tableau 3.2 – Matrice d'expérimentation.

Essai n°	Concentration en or (g/l) <i>Facteur 1</i>	Densité de courant (A/dm ²) <i>Facteur 2</i>	Concentration en cobalt (g/l) <i>Facteur 3</i>	Vitesse (mg/min)	Teneur en cobalt (ppm)
1	2	5	0,5		
2	15	5	0,5		
3	2	25	0,5		
4	15	25	0,5		
5	2	5	1,5		
6	15	5	1,5		
7	2	25	1,5		
8	15	25	1,5		

Tableau 3.3 – Matrice d'expériences.

Essai n°	Concentration en or <i>Facteur 1</i>	Densité de courant <i>Facteur 2</i>	Concentration en cobalt <i>Facteur 3</i>	Vitesse (mg/min)	Teneur en cobalt (ppm)
1	-1	-1	-1		
2	+1	-1	-1		
3	-1	+1	-1		
4	+1	+1	-1		
5	-1	-1	+1		
6	+1	-1	+1		
7	-1	+1	+1		
8	+1	+1	+1		
Niveau -	2 g/l	5 A/dm ²	0,5 g/l		
Niveau +	15 g/l	25 A/dm ²	1,5 g/l		

3.1.3 Expérimentation

Les essais sont menés conformément au plan d'expériences et les résultats sont consignés dans la matrice d'expériences (Tableau 3.4).

Tableau 3.4 – Matrice d'expériences et résultats.

Essai n°	Concentration en or <i>Facteur 1</i>	Densité de courant <i>Facteur 2</i>	Concentration en cobalt <i>Facteur 3</i>	Vitesse (mg/min)	Teneur en cobalt (ppm)
1	-1	-1	-1	53	4 100
2	+1	-1	-1	122	3 510
3	-1	+1	-1	20	3 950
4	+1	+1	-1	125	1 270
5	-1	-1	+1	48	4 870
6	+1	-1	+1	70	2 810
7	-1	+1	+1	68	7 750
8	+1	+1	+1	134	3 580
Niveau -1	2 g/l	5 A/dm ²	0,5 g/l		
Niveau +1	15 g/l	25 A/dm ²	1,5 g/l		

3.1.4 Interprétation des résultats des calculs

Le même modèle postulé, celui des plans factoriels complets 2^3 , sera utilisé pour les deux réponses :

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{23}x_2x_3 + a_{123}x_1x_2x_3 \quad \{3.1\}$$

Dans ce modèle, les lettres ont la signification suivante : y est l'une des réponses, les x_i sont les niveaux des facteurs, les a_i sont les effets des facteurs, les a_{ij} sont les interactions entre deux facteurs ou interactions d'ordre 2 et a_{123} est l'interaction entre les trois facteurs ou interaction d'ordre 3.

Dans un premier temps, les réponses sont étudiées l'une après l'autre.

■ Vitesse de dépôt

On reporte les résultats des vitesses de dépôt sur le domaine d'étude (Figure 3.4). On constate que la vitesse de 80 mg/min est facilement atteinte et qu'il est possible d'aller plus vite puisque la vitesse maximale est de l'ordre de 134 mg/min. On constate aussi que les fortes vitesses sont du côté des fortes concentrations en or.

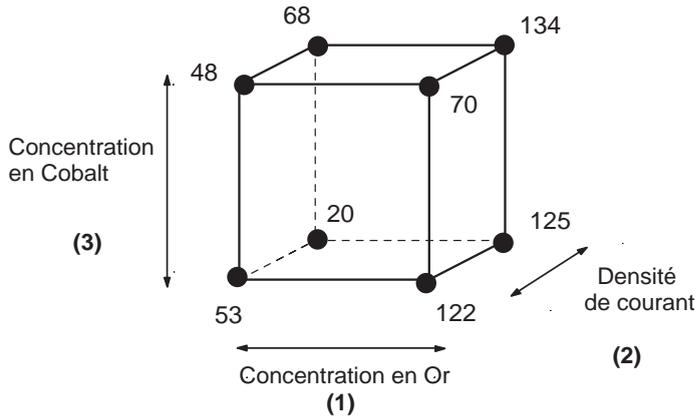


Figure 3.4 – Valeur de la vitesse de dépôt aux sommets du domaine d'étude.

Pour affiner la compréhension des résultats, il faut modéliser les phénomènes. On commence par calculer les effets des facteurs et leurs interactions (Tableau 3.5).

Tableau 3.5 – Effets et interactions des facteurs (unités codées). Réponse « Vitesse de dépôt ».

Effet	Valeur
Constante	80
Concentration en or (1)	32,75
Densité de courant (2)	6,75
Concentration en cobalt (3)	0
Interaction 12	10
Interaction 13	-10,75
Interaction 23	14,25
Interaction 123	1

C'est le facteur 1, la concentration en or, qui est le plus influent. Les facteurs 2 et 3 ne sont pas directement influents mais ils le sont par l'intermédiaire de leurs interactions d'ordre 2. Les trois interactions ont des valeurs à peu près identiques mais de signes différents. L'interaction d'ordre 3 est pratiquement nulle. Ces résultats permettent d'écrire le modèle donnant la vitesse de dépôt en fonction des niveaux des trois facteurs (unités codées) :

$$y_{vitesse} = 80 + 32,75 x_1 + 6,75 x_2 + 10 x_1 x_2 - 10,75 x_1 x_3 + 14,25 x_2 x_3 \quad \{3.2\}$$

Cette relation est utilisée par les logiciels pour faire tous les calculs et tracer tous les graphiques.

On illustre souvent le tableau 3.5 par un diagramme à barres (Figure 3.5) qui permet, d'un coup d'œil, de se rendre compte de l'importance des différents coefficients.

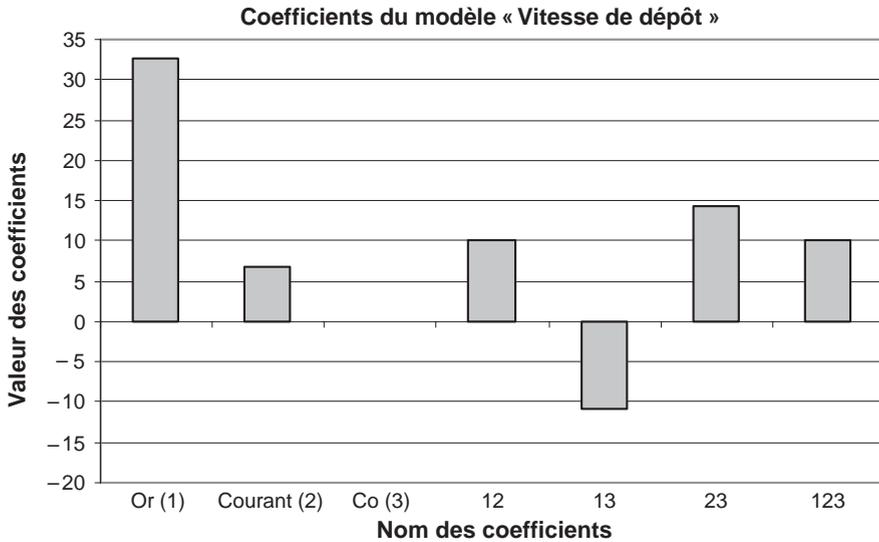


Figure 3.5 – Illustration du tableau 3.5.

Deux autres diagrammes sont utiles pour bien se rendre compte de l'influence des facteurs : le *diagramme des effets* (*Prediction Profiler* dans JMP) qui indique les effets principaux des facteurs et le *diagramme des interactions* (*Interaction Profiles* dans JMP) qui montre les interactions d'ordre 2 entre les facteurs.

□ Diagramme des effets

Nous avons déjà rencontré ce diagramme. Il indique les effets principaux des facteurs, c'est-à-dire les coefficients des termes du premier degré du modèle mathématique. Ce diagramme peut être construit avec les unités codées ou avec les unités normales (Figure 3.6). L'impression visuelle est la même. Pour la présentation des résultats, il est plus commode d'employer les unités normales qui donnent immédiatement les valeurs de comparaison.

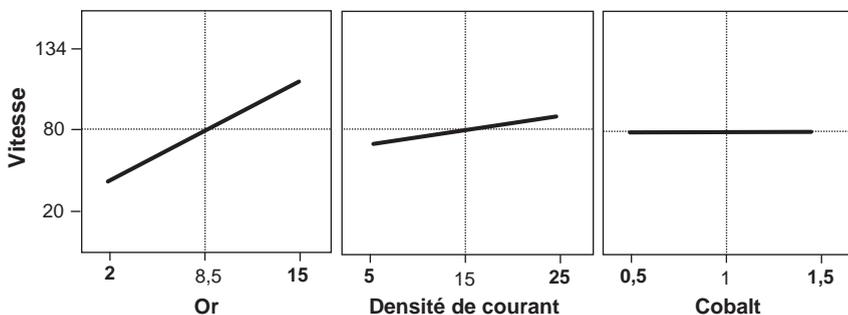


Figure 3.6 – Effet des facteurs.

La vitesse de dépôt est d'autant plus grande que la solution d'électrolyte contient plus d'or et que la densité de courant est plus élevée. La concentration en cobalt de la solution d'électrolyte ne semble jouer aucun rôle. Mais pour avoir une interprétation complète, il faut tenir compte des interactions qui, nous l'avons vu, ne sont pas négligeables.

□ Diagramme des interactions

Le diagramme des d'interactions indique les effets d'un facteur au niveau bas et au niveau haut d'un autre facteur. Le diagramme de la figure 3.7 se lit de la manière suivante : la réponse est indiquée en ordonnées et les échelles des facteurs sont en abscisses. Dans le carré supérieur droit, l'effet du facteur « densité de courant » est indiqué pour les niveaux bas (2) et haut (15) du facteur concentration en Or. Dans le carré inférieur gauche, l'effet du facteur « concentration en or » est indiqué pour les niveaux bas (5) et haut (25) du facteur « densité de courant ».

Nous savons que si les droites des effets ne sont pas parallèles, il y a une interaction significative. C'est le cas de ces deux facteurs dont les pentes des effets sont différentes dans une perspective et se croisent dans l'autre.

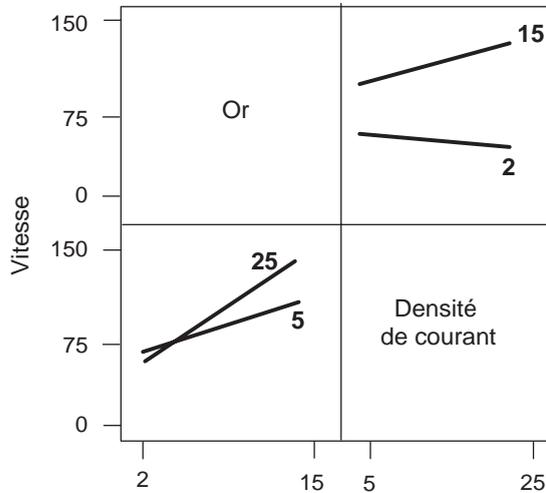


Figure 3.7 – Le diagramme des interactions illustre l'importance des interactions.

On peut présenter le diagramme des interactions en prenant les interactions soit séparément deux à deux, soit toutes regroupées dans un seul tableau (Figure 3.8).

Le diagramme des profils d'interactions montre clairement que les interactions ne sont pas négligeables et qu'il faut les prendre en compte pour faire l'interprétation. La vitesse maximale de dépôt est atteinte lorsque les trois facteurs sont à leur niveau haut : 15 g/l d'or et 1,5 g/l de cobalt dans la solution d'électrolyte et une densité de courant de 25 A/dm² (Figure 3.9).

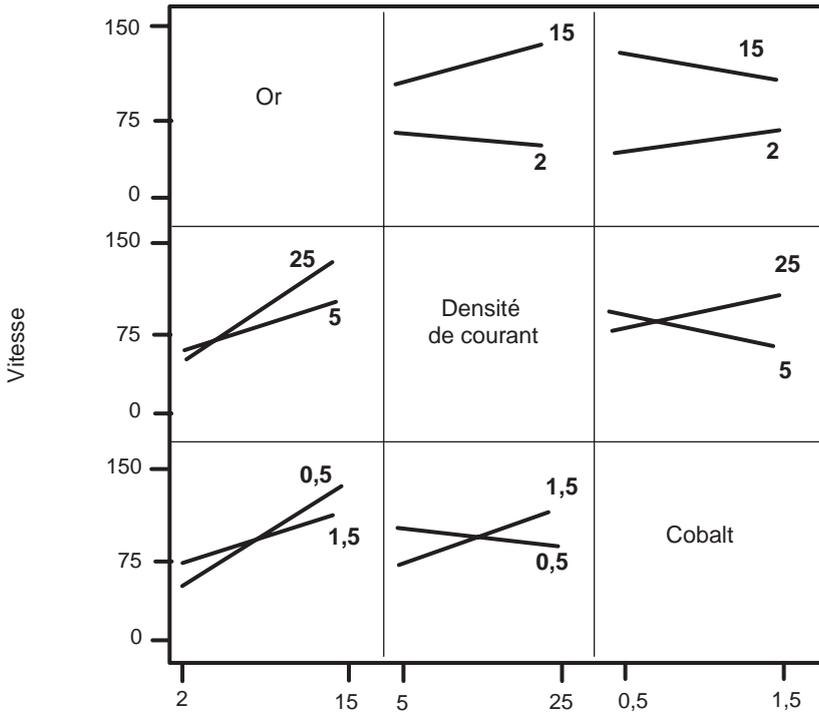


Figure 3.8 – Les profils des interactions sont regroupés en un seul tableau.

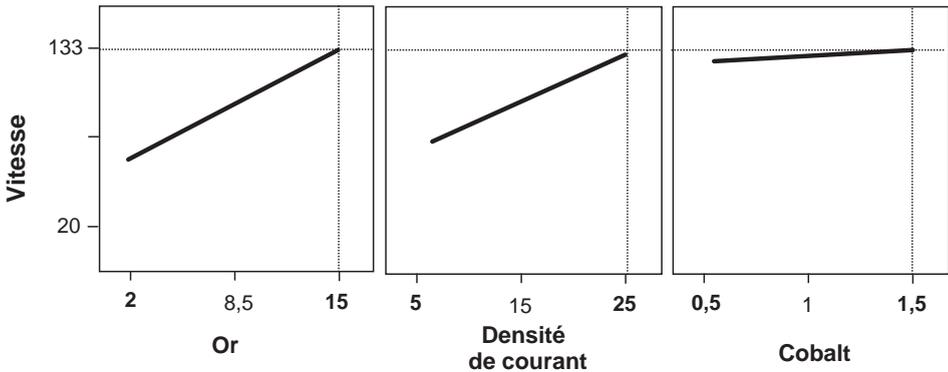


Figure 3.9 – La plus grande vitesse de dépôt est atteinte lorsque les trois facteurs sont à leur niveau haut.

■ Teneur en cobalt de l'alliage

On peut également reporter les résultats des teneurs en cobalt sur le domaine d'étude (Figure 3.10).

On constate que de fortes teneurs en cobalt dans l'alliage sont obtenues lorsque la concentration en or est faible dans la solution d'électrolyte. On constate également

que cette teneur en cobalt augmente avec la densité de courant et avec la concentration en cobalt dans l'électrolyte.

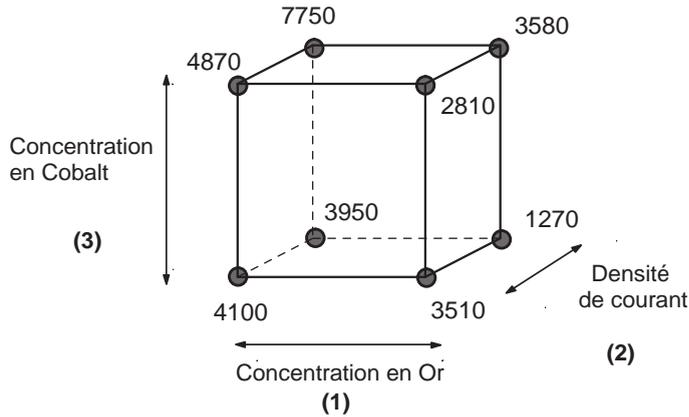


Figure 3.10 – Valeur de la réponse « Teneur en cobalt de l'alliage » aux sommets du domaine d'étude.

Calculons maintenant les effets des facteurs et leurs interactions (Tableau 3.6).

Tableau 3.6 – Effets et interactions des facteurs (unités codées). Réponse « Teneur en cobalt de l'alliage ».

Effet	Valeur
Constante	3 980
Concentration en or (1)	-1 187,5
Densité de courant (2)	157,5
Concentration en cobalt (3)	772,5
Interaction 12	-525
Interaction 13	-370
Interaction 23	755
Interaction 123	-2,5

Ce sont les facteurs 1 (la concentration en or) et 3 (la concentration en cobalt) qui sont les plus influents (Figure 3.11). Mais ils sont de sens opposés. L'effet du facteur 1 est négatif, cela signifie que la réponse baisse quand le facteur passe du niveau bas au niveau haut. L'effet du facteur 3 est positif, cela signifie que la réponse augmente

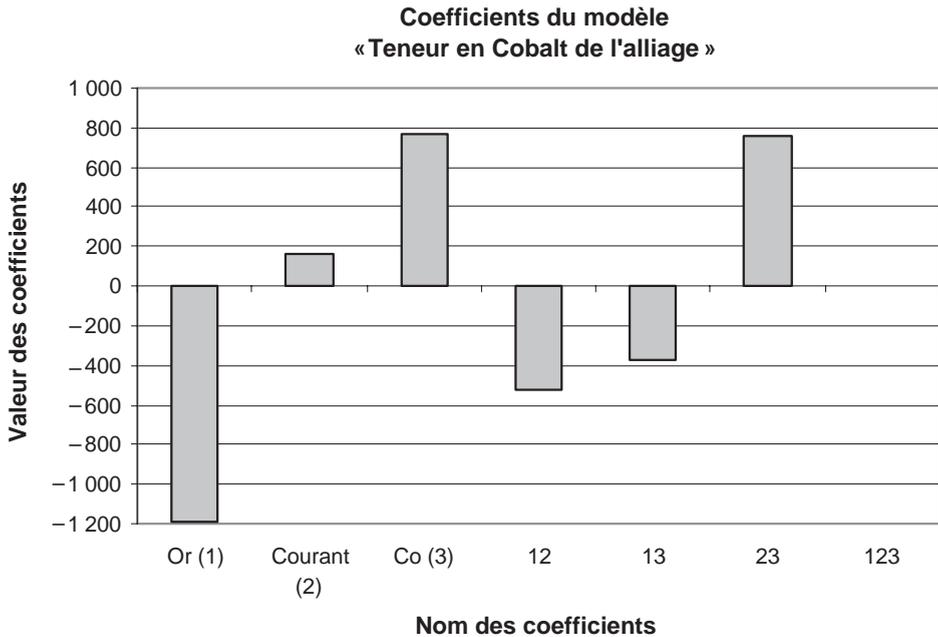


Figure 3.11 – Illustration du tableau 3.6.

quand le facteur passe du niveau bas au niveau haut. Les interactions d'ordre 2 sont importantes (Figure 3.12). L'interaction d'ordre 3 est négligeable.

Ces résultats permettent d'écrire le modèle donnant la concentration en cobalt de l'alliage en fonction des niveaux des trois facteurs (unités codées) :

$$y_{cobalt} = 3\,980 - 1\,187,5 x_1 + 157,5 x_2 + 772,5 x_3 - 525 x_1 x_2 - 370 x_1 x_3 + 755 x_2 x_3 \quad \{3.3\}$$

Cette relation est utilisée par les logiciels pour faire tous les calculs et tracer tous les graphiques.

La teneur en cobalt de l'alliage doit être comprise entre 3 500 et 4 500 ppm. Ces deux valeurs peuvent être atteintes puisque la réponse varie de 1 270 à 7 750 ppm à l'intérieur du domaine d'étude. Pour avoir une vue d'ensemble, il faut tracer les courbes isoréponses de la surface de réponse correspondant aux deux facteurs les plus influents : la concentration en or et la concentration en cobalt (Figure 3.13). La courbe 4 000 ppm a été tracée et les limites de 3 500 et 4 500 ppm ont été indiquées. Les zones grisées sont en dehors de l'objectif et les zones claires rassemblent toutes les combinaisons de concentrations qui permettent d'atteindre l'objectif. Le diagramme a été tracé pour trois niveaux de la densité de courant : 5, 15 et 25 A/dm². Plus la densité de courant est élevée et plus la zone autorisée est étroite.

Ces diagrammes peuvent être construits avec un tableur ou avec un logiciel de plans d'expériences.

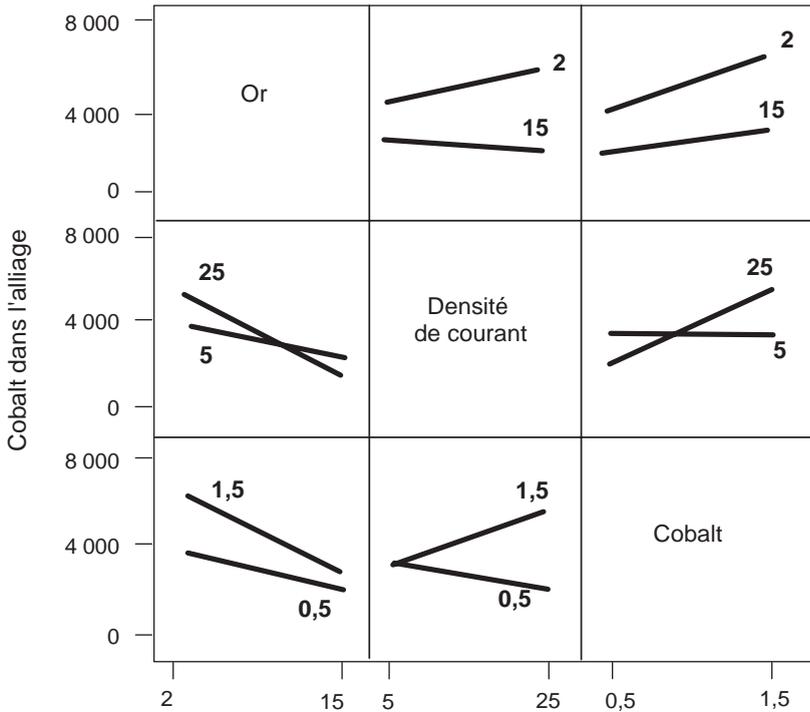


Figure 3.12 – Profils des interactions pour la réponse « Teneur en cobalt de l'alliage ».

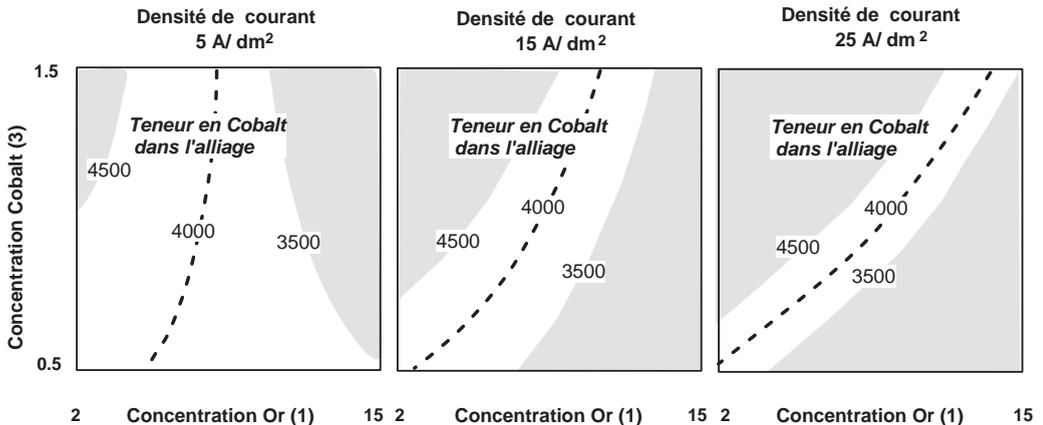


Figure 3.13 – L'objectif de teneur en cobalt n'est pas atteint pour les zones grisées.

■ Recherche de conditions opératoires respectant les objectifs

Comme il y a deux objectifs, une vitesse de dépôt au moins égale à 80 mg/min et une teneur en cobalt de l'alliage comprise entre 3 500 et 4 500 ppm, nous allons superposer les résultats de la vitesse de dépôt à ceux de la teneur en cobalt (Figure 3.14). On interdit les vitesses inférieures à 80 mg/min.

Pour les faibles densités de courant on constate que la zone des faibles concentrations en or est supprimée. Il ne reste qu'une zone utile de fortes concentrations en or et de faibles concentrations en cobalt.

Pour les densités de courant moyennes, on constate également que la zone des faibles concentrations en or est supprimée. Il ne reste qu'une zone où les deux objectifs sont respectés : elle se situe aux fortes concentrations en or et aux fortes concentrations en cobalt.

Pour les fortes densités de courant, la zone des faibles concentrations en or est supprimée. Il ne reste qu'une zone qui ressemble à celle des densités moyennes de courant.

Toutes les combinaisons appartenant aux zones non grisées permettent de répondre aux objectifs. L'expérimentateur a donc un large éventail de solutions. Il pourra choisir une solution répondant à une contrainte supplémentaire, par exemple une solution donnant un meilleur aspect de surface ou une solution plus économique.

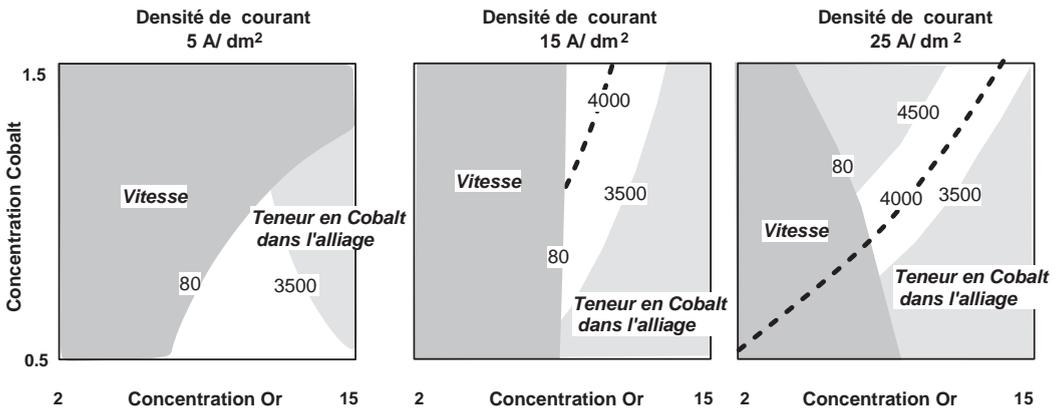


Figure 3.14 – Les objectifs sont respectés dans les zones non grisées.

■ Peut-on faire mieux ?

On peut se poser la question de savoir dans quelles conditions on peut travailler un peu plus vite, par exemple à 100 mg/min. Il suffit de remplacer la valeur de la limite de vitesse de 80 mg/min par 100 mg/min. On obtient des diagrammes analogues à ceux de la figure 3.15.

Si l'on veut travailler à 100 mg/min, il faut choisir une densité de courant soit faible soit forte. Il n'y a aucune solution si la densité de courant est moyenne.

Pour les faibles densités de courant, il faut une forte concentration en or et une faible concentration en cobalt.

Pour les fortes densités de courant, il faut une forte concentration en or et une forte concentration en cobalt.

Grâce aux modèles établis, l'expérimentateur peut calculer (avec le logiciel) des prévisions de réponses pour des conditions opératoires données. Par exemple, il peut décider de privilégier les solutions faisant intervenir les fortes densités de courant qui permettent d'atteindre une concentration en cobalt de 4 000 ppm

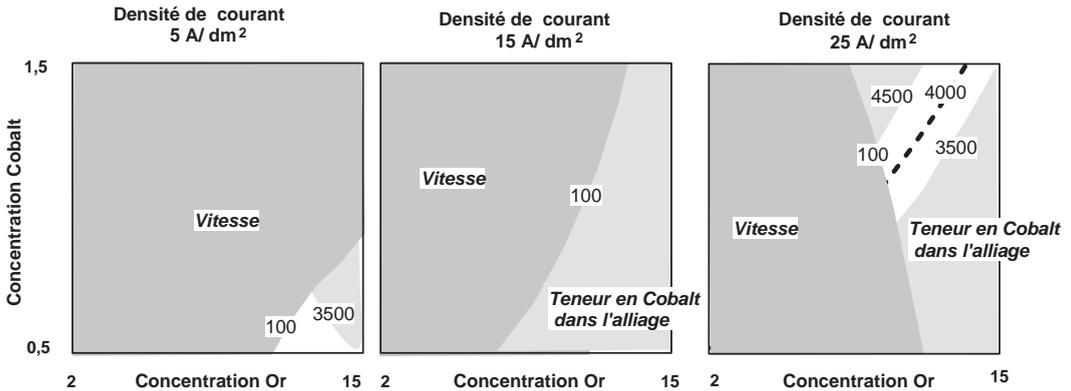


Figure 3.15 – Avec une vitesse de dépôt de 100 mg/min, les objectifs sont respectés uniquement dans deux petites zones non grisées.

alors que les faibles densités de courant ne le permettent pas. Dans cette hypothèse, il ne peut pas conclure sans effectuer une vérification. Il va donc prévoir une ou plusieurs expériences de contrôle (Figure 3.16). L'expérimentateur ajuste les facteurs fixés à leurs niveaux d'expérimentation et il attribue les valeurs suivantes aux facteurs du plan :

- La concentration en or de la solution d'électrolyte est réglée à 12 g/l.
- La concentration en cobalt de la solution d'électrolyte est réglée à 1,3 g/l.
- La densité de courant est réglée à 25 A/dm².

Avec les modèles établis, l'expérimentateur calcule que la vitesse devrait être de 115 mg/min et la teneur en cobalt de l'alliage devrait être de 4 011 ppm.

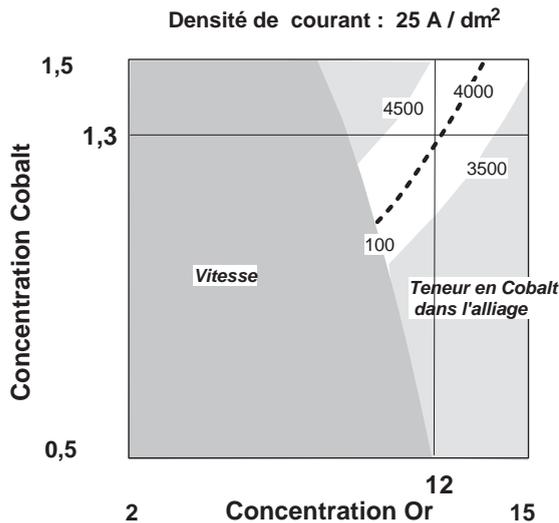


Figure 3.16 – Les prévisions de réponses indiquent que les objectifs pourraient être atteints si la concentration en or de l'électrolyte était de 12 mg/l et celle du cobalt de 1,3 mg/l pour une densité de courant de 25 A/dm².

Il réalise deux essais de contrôle en reprenant ces conditions opératoires. Il trouve une vitesse de dépôt de 110 et 105 mg/min et une teneur en cobalt dans l'alliage de 3 950 et 3 920 ppm. Ces résultats le satisfont car ils confirment ses prévisions. L'expérimentateur va pouvoir faire ses recommandations.

3.2 Conclusion de l'étude

L'expérimentateur précise les conditions opératoires permettant de réaliser des dépôts d'or fin répondant aux impératifs de qualité définis au départ de l'étude. Il indique : le matériel ; la nature, l'origine et les propriétés des matières premières ; la mise en œuvre des matières premières ; les niveaux des facteurs qui ont été fixés pendant l'expérimentation. Enfin, il rappelle les réglages des quatre principaux facteurs :

- le pH de l'électrolyte réglé à 4,5 ;
- la concentration en or de la solution d'électrolyte réglée à 12 g/l ;
- la concentration en cobalt de la solution d'électrolyte réglée à 1,3 g/l ;
- la densité de courant réglée à 25 A/dm².

Dans ces conditions, on doit obtenir une vitesse de dépôt légèrement supérieure à 100 mg/min et une teneur en cobalt dans l'alliage de l'ordre 3 900 ppm. Les objectifs de l'étude sont respectés et même un peu dépassés.

4 • PLANS FACTORIELS COMPLETS À 4 FACTEURS

Nous avons vu un plan factoriel complet à deux facteurs et un autre à trois facteurs. Dans ces deux plans les facteurs ne prenaient que deux niveaux, un niveau bas et un niveau haut. Le premier plan a été désigné comme plan 2^2 (deux facteurs prenant chacun deux niveaux). Le second plan a été désigné comme plan 2^3 (trois facteurs prenant chacun deux niveaux). Nous allons voir maintenant un plan faisant intervenir quatre facteurs prenant chacun deux niveaux, c'est-à-dire un plan 2^4 . Le nombre d'essais est donc de 16, ce qui est encore raisonnable. Plus le nombre de facteurs est élevé, plus le nombre d'essais devient grand. Ce nombre d'essais augmente exponentiellement avec le nombre de facteurs. Pour 5 facteurs, il faudrait réaliser 32 essais et, pour 6 facteurs, 64 essais. Il est donc rare d'exécuter un plan factoriel complet ayant plus de 5 ou 6 facteurs. Nous verrons qu'il est possible, moyennant certaines hypothèses, de réduire le nombre d'essais sans diminuer le nombre de facteurs. Les plans factoriels fractionnaires ont été inventés pour résoudre ce problème.

Dans le présent chapitre nous traitons l'exemple de la galette des rois. Cet exemple ne pose aucun problème de construction. Mais nous verrons que l'interprétation n'est pas automatique et que l'expérimentateur doit faire preuve de connaissances et savoir faire appel à toutes les compétences pour choisir la bonne décision. L'expérimentateur est celui qui sait présenter les différentes alternatives aux décideurs pour que l'entreprise fasse le bon choix.

Il existe en France une tradition qui s'appelle la « galette des rois ». À la fin du mois de décembre et pendant tout le mois de janvier, on trouve, chez les détaillants, des gâteaux particuliers appelés « galettes des rois ». Ce gâteau présente la particularité d'avoir un petit objet, la fève, glissé et dissimulé dans la pâte. Cette galette est dégustée en famille ou avec des amis. Le Papa ou la Maman découpe la galette en un nombre de parts exactement égal au nombre de convives. Le plus jeune des enfants se place de manière à ne pas voir la galette, par exemple dans une pièce voisine ou sous la table. La personne qui a découpé la galette désigne l'une des parts et demande à l'enfant : « Pour qui cette part-là ? » L'enfant répond, par exemple : « Pour ma sœur. » La part est attribuée à la sœur. On poursuit ainsi la distribution jusqu'à ce que chacun ait sa part. Celui qui trouve la fève dans sa part de galette est le roi, ou la reine, de la soirée, il est applaudi et congratulé par tous. Les enfants adorent la galette des rois et espèrent bien devenir le roi ou la reine de la soirée... mais je crois que c'est pareil pour les grandes personnes. On notera que

les petits Français sont bien entraînés pour réaliser une distribution parfaitement aléatoire et que la galette des rois est une bonne initiation à la science statistique. Cet exemple sera l'occasion de présenter une méthode d'optimisation multicritère très utile : la *désirabilité*.

4.1 Exemple 03 : la galette des rois

4.1.1 Préparation du plan d'expériences

■ Description de l'étude

La galette des rois est préparée à partir d'une pâte feuilletée spéciale que l'on étale en plaque d'environ 1 centimètre d'épaisseur et que l'on découpe en forme de cercle. Ces cercles de pâte sont appelés des abaisses. Pour réaliser une galette il faut deux abaisses. Sur la première abaisse, l'abaisse inférieure, on dépose un fourrage de pâte d'amandes et l'on place la fève de manière à ce qu'elle soit parfaitement dissimulée. On recouvre l'ensemble d'une deuxième abaisse identique à la première. On perce l'abaisse supérieure de nombreux petits trous pour que la vapeur d'eau puisse s'échapper pendant la cuisson. On obtient une galette que l'on introduit dans un four chauffé de manière homogène à une certaine température et que l'on laisse cuire pendant 30 minutes. Pour la présente étude il s'agit de galettes industrielles qui, une fois cuites sont congelées et acheminées dans les magasins grâce à un réseau de distribution. À la sortie du four les galettes sont examinées et seules celles qui ne présentent aucun défaut sont congelées. Les galettes sont ensuite conditionnées et distribuées chez les revendeurs en préservant la chaîne du froid. Les galettes sont vendues en général dans les trois semaines qui suivent leur fabrication.

Tout serait pour le mieux si des clients mécontents n'étaient pas revenus réclamer en faisant constater que certaines galettes avaient des abaisses décollées et présentaient des fentes sur l'abaisse inférieure. Le principal grief était que les galettes laissaient apparaître la fève et que la magie de la galette des rois s'évanouissait alors complètement.

Devant une telle catastrophe, l'industriel réagit immédiatement en convoquant ses meilleurs techniciens. La situation est analysée et la décision de faire une étude pour trouver l'origine des deux défauts est prise sans tarder. Les galettes ne doivent présenter aucun défaut 4 semaines après la sortie d'usine. Les défauts visés sont le décollement et les fentes, qui en jargon technique de pâtisserie, s'appellent des fêles.

■ Objectif de l'étude

L'objectif de l'étude est de trouver les conditions de fabrication qui permettront de réduire, et si possible d'éliminer, les décollements et les fêles sur les galettes congelées ayant 4 semaines de vie.

■ Réponses

Le choix de la réponse permettant de bien mesurer la grandeur d'intérêt et de corriger les défauts constatés est bien souvent difficile à définir. Ici, la mesure du décollement se heurte à plusieurs difficultés. Le décollement n'est pas régulier et l'on a constaté

plusieurs zones de décollement plus ou moins longues sur la même galette. Il peut y avoir des parties bien collées, d'autres à moitié collées et enfin d'autres complètement décollées. Les tentatives d'évaluation du décollement par des mesures de longueur, de profondeur et d'épaisseur se soldèrent par des échecs. Il en fut de même pour la mesure des fêles, celles-ci étant plus ou moins larges, plus ou moins profondes et plus ou moins rectilignes. On finit donc par établir un barème avec des défauts tests, ce qui permit de donner des notes à toutes les galettes participant à l'étude.

Le décollement fut estimé par une note comprise entre 0 et 3 :

- 3 est l'absence de décollement ;
- 0 est le plus mauvais décollement constaté.

La fêle fut estimée par une note comprise entre 0 et 3 :

- 3 est l'absence de fêle ;
- 0 est la galette possédant le plus grand nombre de fêles.

L'objectif chiffré est donc 3 pour le décollement et 3 pour la fêle.

■ Facteurs et domaine d'étude

L'industriel et son équipe établirent la liste de tous les facteurs pouvant influencer l'apparition des défauts. Puis, ils sélectionnèrent ceux qui leur paraissaient susceptibles d'engendrer les deux défauts :

- Facteur 1 : rapport du poids des abaisses. C'est le rapport : poids de l'abaisse supérieure/poids de l'abaisse inférieure :
 - niveau bas : 0,9
 - niveau haut : 1,1
- Facteur 2 : nombre de trous dans l'abaisse supérieure :
 - niveau bas : peu
 - niveau haut : beaucoup
- Facteur 3 : profil de chauffage du four de cuisson :
 - niveau bas : fort au début
 - niveau haut : fort à la fin
- Facteur 4 : durée de la cuisson :
 - niveau bas : 29 min
 - niveau haut : 31 min

Les facteurs et leurs niveaux sont rassemblés dans le tableau 4.1.

Tableau 4.1 – Facteurs retenus et domaine d'étude.

Facteur	Niveau bas	Niveau moyen	Niveau haut
Rapport du poids des abaisses (1)	0,90	1	1,1
Nombre de trous (2)	Peu	Normal	Beaucoup
Profil de chauffage (3)	Fort au début	Homogène	Fort à la fin
Durée de la cuisson (4)	29 min	30 min	31 min

Le niveau moyen correspond aux conditions de la fabrication utilisées avant le plan d'expériences.

■ Choix du plan d'expériences

Le responsable de la fabrication décide d'étudier les 4 facteurs avec un plan factoriel complet 2^4 , car il s'attend à de nombreuses interactions entre les facteurs. Le plan comporte donc 16 essais (Tableau 4.2).

Tableau 4.2 – Plan d'expériences pour l'étude sur la galette des rois.

Essai n°	Rapport du poids des abaisses (1)	Nombre de trous (2)	Chauffage du four (3)	Durée de la cuisson (4)
1	-	-	-	-
2	+	-	-	-
3	-	+	-	-
4	+	+	-	-
5	-	-	+	-
6	+	-	+	-
7	-	+	+	-
8	+	+	+	-
9	-	-	-	+
10	+	-	-	+
11	-	+	-	+
12	+	+	-	+
13	-	-	+	+
14	+	-	+	+
15	-	+	+	+
16	+	+	+	+
-1	0,90	Peu	Fort au début	29 min
+1	1,1	Beaucoup	Fort à la fin	31 min

La présentation du plan suit l'ordre classique des essais ou ordre de Yates. Les niveaux de la première colonne sont alternés et commencent par un -1 . Les niveaux de la deuxième colonne sont alternés deux par deux : 2 signes -1 et 2 signes $+1$ et ainsi de suite. Les niveaux de la troisième colonne sont alternés quatre par quatre : 4 signes -1 et 4 signes $+1$ et ainsi de suite. Les niveaux de la quatrième colonne sont alternés huit par huit : 8 signes -1 et 8 signes $+1$. Cet ordre sera utilisé pour présenter tous les plans d'expériences du livre et tous les fichiers informatiques. Mais l'ordre d'exécution des essais d'un plan n'est pas forcément l'ordre de présentation. L'ordre des essais est une question importante que nous étudierons en détail au cours du chapitre 8.

4.1.2 Expérimentation

Les galettes sont cuites dans un four industriel continu. À l'entrée du four, elles sont déposées sur un tapis roulant et disposées les unes derrière les autres sur plusieurs rangées. Elles avancent régulièrement dans le four et cuisent progressivement, elles y séjournent le temps imposé par le plan d'expériences. Elles sont récupérées à la sortie du four, puis refroidies. On sélectionne celles qui ne présentent aucun défaut et on les congèle. On prélève, au hasard, un échantillon de 100 galettes préparées dans les mêmes conditions. Elles sont emballées, étiquetées et stockées pendant 4 semaines. Au bout de quatre semaines, chaque galette passe entre les mains d'un juge qui lui attribue une note pour la fêlé et une autre pour le décollement. Un second juge vérifie les notes du premier et, en cas de désaccord, les deux juges se concertent pour mettre la note la plus précise possible. Pour chaque essai du plan on lance une fabrication particulière telle que l'on vient de la décrire.

Un tel plan nécessite une excellente organisation pour être mené à bien et l'enjeu doit être important car le coût d'une telle recherche est considérable. Heureusement, dans notre cas tous les essais et toutes les mesures ont pu être réalisés. Nous ne donnerons pas les résultats des 100 galettes par essais. Nous nous contenterons des moyennes pour chacune des deux réponses. Les résultats figurent dans le tableau 4.3.

Tableau 4.3 – Résultats du plan d'expériences « Galette des rois ».

Essai n°	Poids des abaisses (1)	Nombre de trous (2)	Chauffage (3)	Durée de cuisson (4)	Décollement	Fêlé
1	-	-	-	-	2,8	2,2
2	+	-	-	-	3	2,8
3	-	+	-	-	2,2	1,2
4	+	+	-	-	3	2,6
5	-	-	+	-	3	1,4
6	+	-	+	-	2,6	2
7	-	+	+	-	3	1,4

Tableau 4.3 – Résultats du plan d'expériences « Galette des rois ». (Suite)

Essai n°	Poids des abaisses (1)	Nombre de trous (2)	Chauffage (3)	Durée de cuisson (4)	Décollement	Fêle
8	+	+	+	–	2,4	2,2
9	–	–	–	+	1,4	2,4
10	+	–	–	+	2,8	3
11	–	+	–	+	0,8	1,2
12	+	+	–	+	1,6	3
13	–	–	+	+	1,4	1,4
14	+	–	+	+	2,6	3
15	–	+	+	+	3	1,8
16	+	+	+	+	2,8	3
–1	0,90	Peu	Fort au début	29 min		
+1	1,1	Beaucoup	Fort à la fin	31 min		

C'est à partir de ce tableau que l'interprétation peut débuter.

4.1.3 Interprétation des résultats du plan

■ Modélisation mathématique

Le modèle mathématique choisi *a priori* (modèle PDAI) est le suivant :

$$\begin{aligned}
 y = & a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 \\
 & + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{14}x_1x_4 + a_{23}x_2x_3 + a_{24}x_2x_4 + a_{34}x_3x_4 \\
 & + a_{123}x_1x_2x_3 + a_{124}x_1x_2x_4 + a_{134}x_1x_3x_4 + a_{234}x_2x_3x_4 \\
 & + a_{1234}x_1x_2x_3x_4
 \end{aligned}$$

Ce modèle mathématique comporte :

- un coefficient constant a_0 ;
- quatre termes comprenant un coefficient a_i (effets des facteurs) multiplié par les niveaux respectifs de chaque facteur ;
- des termes comprenant un coefficient a_{ij} (interactions entre deux facteurs), et le produit des niveaux des facteurs pris deux à deux ;
- des termes comprenant un coefficient a_{ijk} (interactions entre trois facteurs), et le produit des niveaux des facteurs pris trois à trois ;
- un terme comprenant un coefficient a_{ijkl} (interaction entre quatre facteurs), et le produit des niveaux des facteurs pris quatre à quatre.

Il y a 16 coefficients, soit 16 inconnues. Les 16 essais vont permettre de calculer ces 16 inconnues. La résolution de ce système d'équations est réalisée grâce au logiciel.

Le même modèle de base sera utilisé pour les deux réponses. Mais les coefficients seront fort probablement différents d'une réponse à l'autre. La représentation mathématique finale ne sera pas la même pour chacune des réponses.

■ Préparation des calculs

Quel que soit le logiciel que vous utilisez, vous devez entrer les résultats expérimentaux dans l'ordinateur. Vous prendrez bien garde d'associer correctement chaque résultat avec l'essai correspondant du plan d'expériences. Un décalage ou une erreur de transcription ruinerait les efforts de recherche et pourrait conduire à une interprétation erronée des résultats.

Pour le détail des manipulations permettant de lancer les calculs avec le logiciel, reportez-vous au chapitre 14.

■ Examiner les résultats des calculs

Les résultats des calculs sont affichés dans le tableau 4.4.

Tableau 4.4 – Effets et interactions des facteurs (unités codées).

Effet	Décollement	Fêle
Moyenne	2,4	2,16
Poids des abaisses (1)	0,2	0,54
Trous (2)	-0,05	-0,11
Chauffage (3)	0,2	-0,14
Durée de cuisson (4)	-0,35	0,19
Interaction 12	-0,1	0,11
Interaction 13	-0,2	-0,01
Interaction 14	0,2	0,11
Interaction 23	0,25	0,19
Interaction 24	0,05	0,01
Interaction 34	0,2	0,09
Interaction 123	-0,1	-0,14
Interaction 124	-0,15	-0,01
Interaction 134	0,05	0,06
Interaction 234	0,2	0,01
Interaction 1234	0	-0,06

Trois facteurs sont influents sur le décollement :

- Facteur 1 : rapport du poids des abaisses.
- Facteur 3 : profil de chauffage du four de cuisson.
- Facteur 4 : durée de la cuisson.

Il y a de nombreuses interactions d'ordre 2 et d'ordre 3 qui sont du même ordre de grandeur que les effets principaux. Cela n'a rien de surprenant dans le domaine de la pâtisserie. La figure 4.1 illustre les valeurs des coefficients du modèle de la réponse « Décollement ».

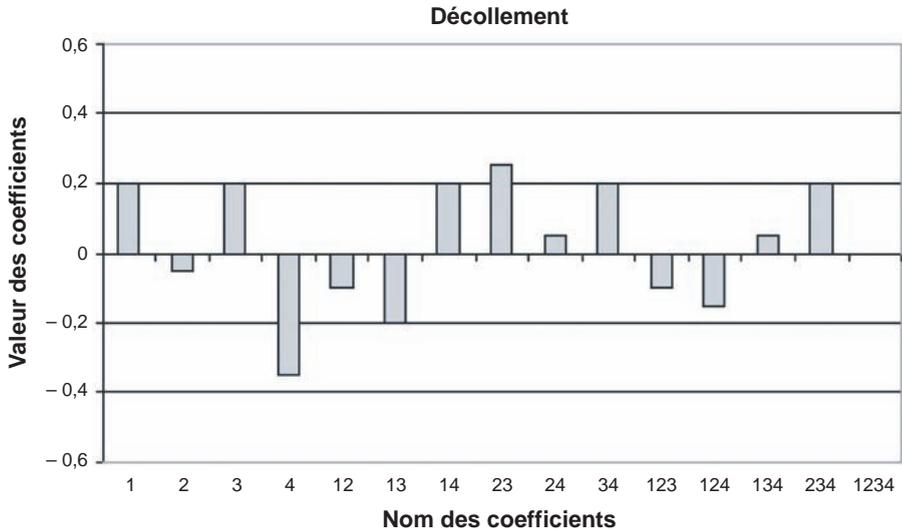


Figure 4.1 – Diagramme à barres des coefficients du modèle de la réponse « Décollement ».

Pour la fêlé, il n'y a qu'un facteur (le rapport du poids des abaisses) nettement plus influent que les trois autres. Mais aucun ne peut être négligé car tous les facteurs modifient suffisamment la note pour être pris en compte. On vérifie qu'il y a aussi de nombreuses interactions d'ordre 2 et d'ordre 3. La figure 4.2 illustre le tableau 4.4.

L'objectif de l'étude est d'améliorer les deux réponses en même temps pour diminuer le décollement et la fêlé. On cherche les conditions qui permettent d'atteindre une note de 3 pour les deux réponses.

■ Interpréter les résultats des calculs

La présence des interactions ne permet pas d'étudier l'influence d'un facteur à partir du seul effet principal sur une réponse car lorsqu'on change le niveau d'un facteur, on modifie les effets de tous les autres facteurs. Au centre du domaine d'étude, la note du décollement est 2,4 et la note de la fêlé est 2,1625 (Figure 4.3). Une interprétation basée uniquement sur les effets principaux des facteurs serait source d'erreurs, il faut tenir compte des interactions. On va progresser par étapes en tenant compte du sens de variation des effets pour les deux réponses. Le calcul des réponses en fonction du niveau des facteurs est effectué grâce à un logiciel de plans d'expériences. Avec le logiciel JMP, c'est l'outil **Prediction Profiler** qu'il faut utiliser.

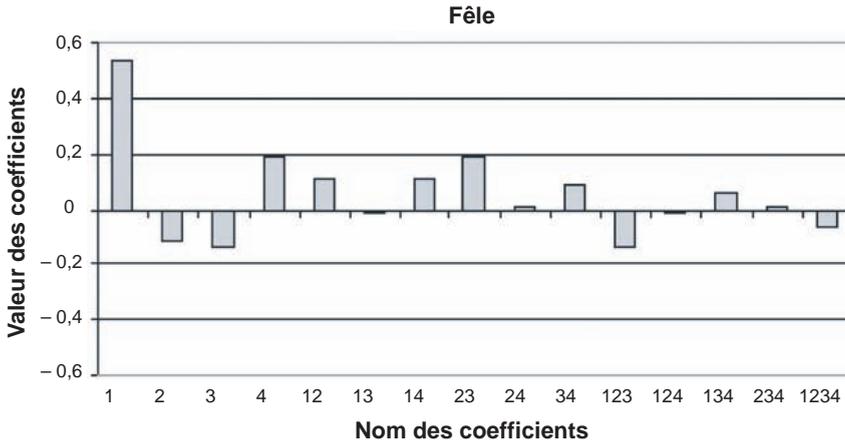


Figure 4.2 – Diagramme à barres des coefficients du modèle de la réponse Fêle.

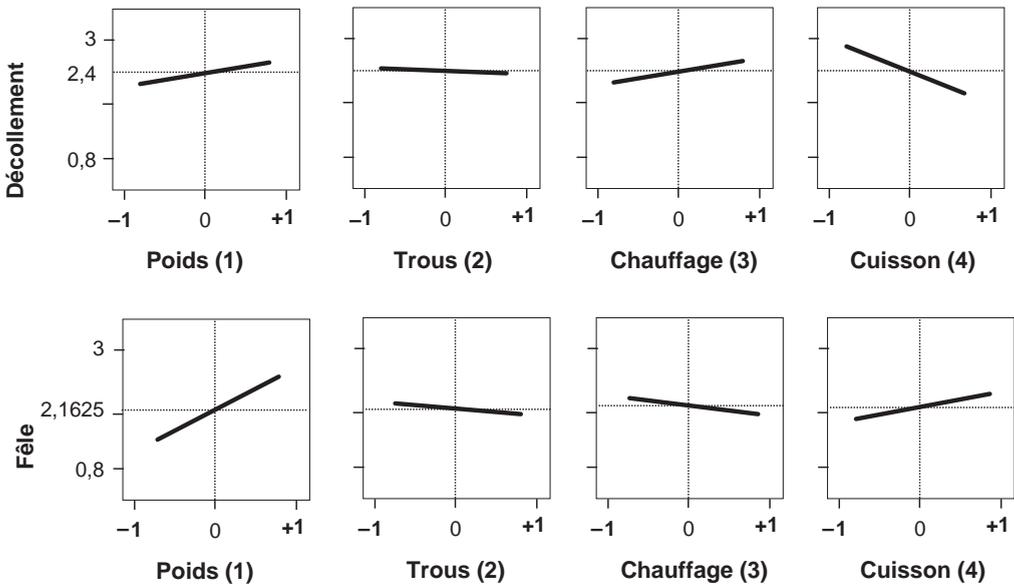


Figure 4.3 – Diagramme des effets des quatre facteurs pour les deux réponses.

Le sens de variation du facteur 1 (le rapport du poids des abaisses) est le même pour les deux réponses. Donc, quand on améliore (détériore) l'une, on améliore (détériore) l'autre. Cela est intéressant puisque c'est le seul facteur sur lequel on peut jouer pour améliorer la fêle. On obtiendra, en même temps, des galettes avec moins de fêle et moins de décollement. On règle donc le facteur 1 au niveau haut (Figure 4.4). La note du décollement passe à 2,6 et celle de la fêle à 2,7.

Pour cette valeur du niveau du facteur 1, on s'aperçoit que le facteur 2 ne modifie que le décollement et qu'il est sans effet sur la fêle.

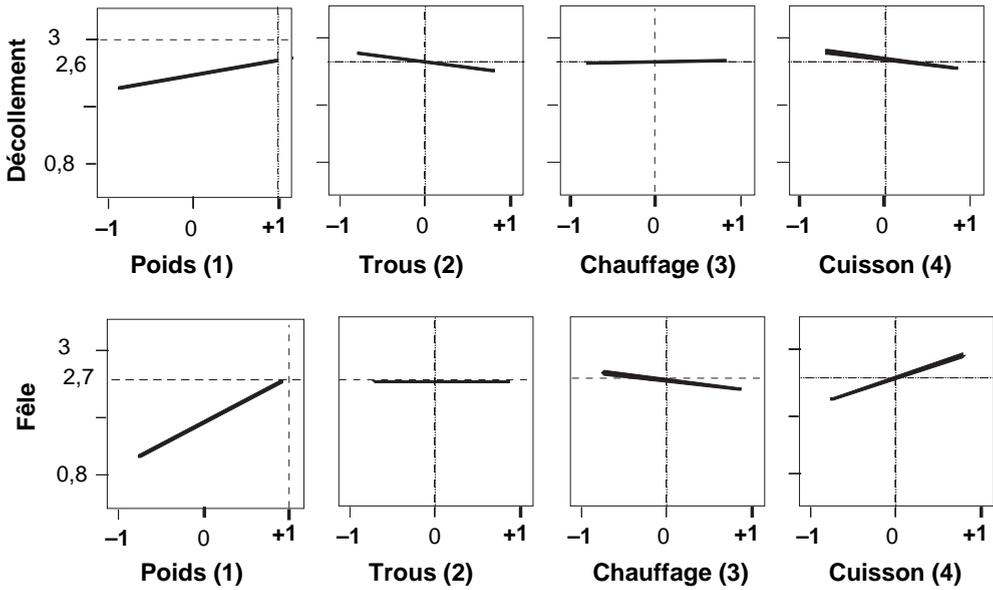


Figure 4.4 – Le facteur 1 est réglé au niveau haut.

On règle ce facteur au niveau bas pour réduire le décollement (Figure 4.5). On remarque que, maintenant, le facteur 3 agit de la même façon sur les deux réponses. Les changements de pente des facteurs sont dus aux interactions.

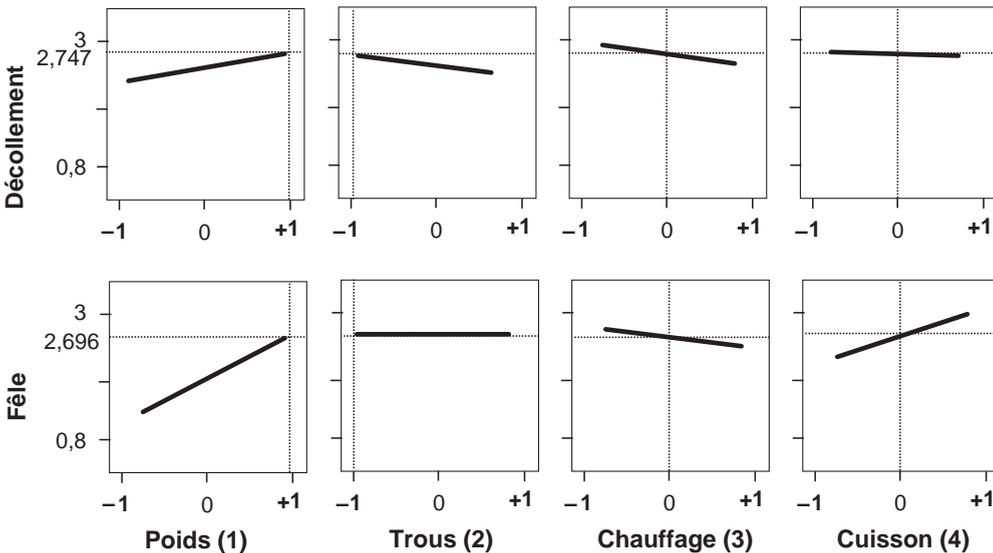


Figure 4.5 – Le facteur 2 est réglé au niveau bas

On profite de cette situation favorable qui évite un compromis : le facteur 3 est placé au niveau bas pour les deux réponses (Figure 4.6).

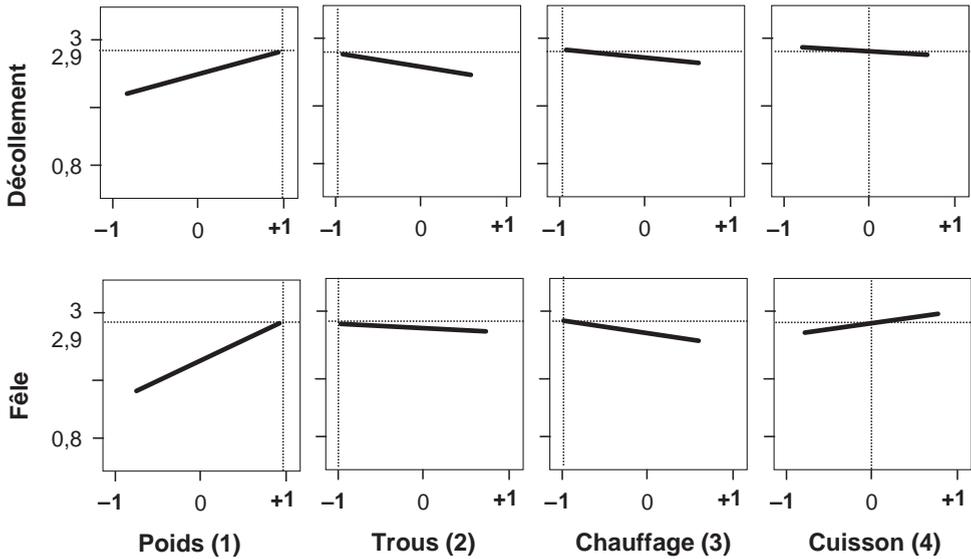


Figure 4.6 – Le facteur 3 est réglé au niveau bas.

Le réglage du facteur 4, la durée de la cuisson, est plus délicat. En effet, ce facteur agit de manière opposée sur les deux réponses. Si on le place au niveau haut, on obtient bien la note 3 pour la fêle, mais on redescend à 2,8 pour le décolllement. Inversement, si on le place au niveau bas, on obtient la note 3 pour le décolllement, mais on redescend à 2,8 pour la fêle. Lorsqu'on laisse le facteur 3 au niveau zéro, on a la note 2,9 pour les deux réponses (Figure 4.7).

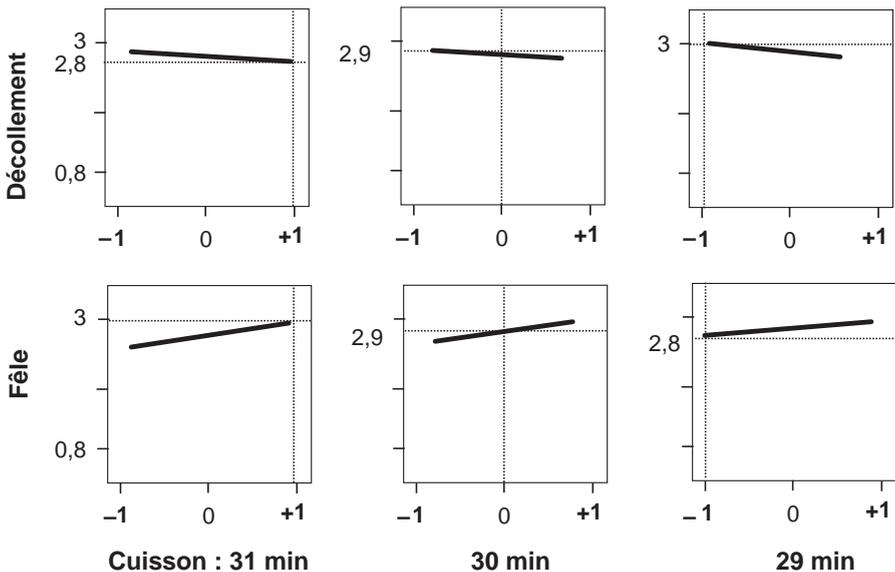


Figure 4.7 – Le facteur 4 peut être réglé à différents niveaux selon la solution recherchée.

Faut-il laisser une durée de cuisson de 30 minutes ou faut-il la réduire pour éliminer les décollements ou l'augmenter pour supprimer les fêles ? La réponse n'appartient pas, normalement, au technicien mais au responsable du marketing qui devra apprécier la moins mauvaise solution étant donné sa clientèle. Les graphiques permettent de présenter les différentes facettes du problème et facilitent les prises de décision. Le technicien possède toutes les données et tous les moyens de les présenter aux décideurs. Le technicien peut également suggérer un complément d'étude pour obtenir la note 3 pour les deux réponses. Lorsqu'on examine la figure 4.6, on constate que l'on peut améliorer la situation :

- en augmentant le rapport du poids des abaisses ;
- en diminuant le nombre de trous ;
- en forçant la température du début de chauffe dans le four.

Si l'enjeu en vaut la peine, l'expérimentateur n'aura aucun mal à obtenir un complément de budget pour finaliser l'étude.

4.2 Fonction désirabilité

La fonction *désirabilité* est très utile quand il faut trouver le meilleur compromis entre plusieurs réponses. Cette fonction a été proposée par Derringer et Suich et elle figure dans de nombreux logiciels de plans d'expériences. Elle est très facile à utiliser et rend de grands services. Nous aurons l'occasion de l'utiliser sur plusieurs exemples.

4.2.1 La fonction désirabilité appliquée à une seule réponse

Commençons par définir la fonction désirabilité pour une seule réponse. Cette fonction varie de 0 à 1. La valeur 1 correspond à une satisfaction maximale, la valeur 0 à un refus complet.

■ Fortes valeurs souhaitées

Par exemple, supposons que vous désiriez un décollement supérieur à 2,5. En deçà de 2,5, vous estimez qu'il y a trop de décollement et donc la fonction désirabilité correspondante vaut 0 pour tous les décollements compris entre 0 et 2,5. En revanche, au-delà de 2,5, vous êtes satisfait et la fonction désirabilité correspondante vaut 1 pour tous les décollements égaux ou supérieurs à 2,5.

On peut représenter la fonction désirabilité comme sur la figure 4.8. Cette fonction reste à la valeur 0 tant que la valeur de la réponse ne convient pas et passe à 1 dès que la valeur de la réponse convient. Mais, cette représentation est brutale et sans nuances. Si vous avez un décollement de 2,49 c'est presque la même chose que 2,51, et pourtant la fonction désirabilité passe de 0 à 1. Il faut introduire un peu de souplesse et de progressivité. Si le décollement atteint 2,6, c'est mieux que 2,4, mais c'est moins bien que 2,7. On va donc introduire des valeurs intermédiaires entre les valeurs qui ne conviennent pas et celles que l'on souhaite. Par exemple, on fixe la grille suivante :

- désirabilité = 0 pour un décollement de 2,3 et moins ;
- désirabilité = 0,25 pour un décollement de 2,4 ;
- désirabilité = 0,50 pour un décollement de 2,5 ;

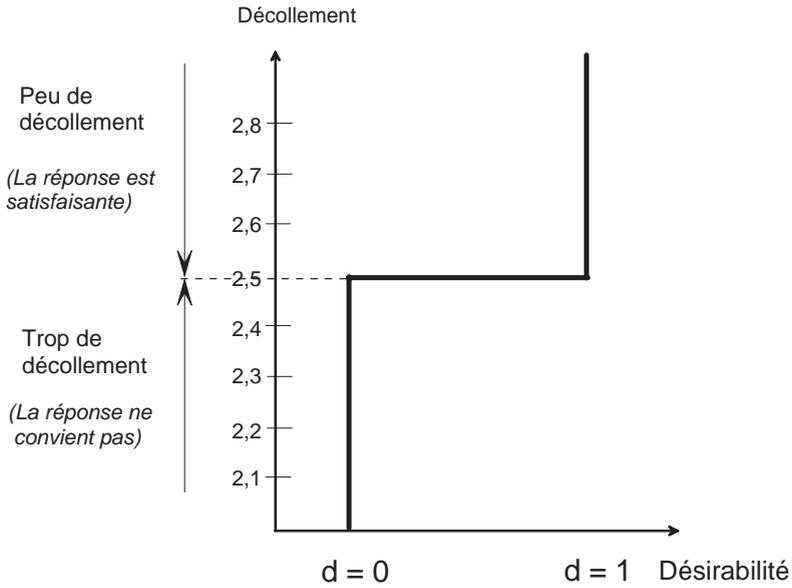


Figure 4.8 – La fonction désirabilité vaut 1 quand la réponse convient. Elle vaut 0 quand la réponse ne convient pas.

- désirabilité = 0,75 pour un décollement de 2,6 ;
- désirabilité = 1 pour un décollement de 2,7 et plus.

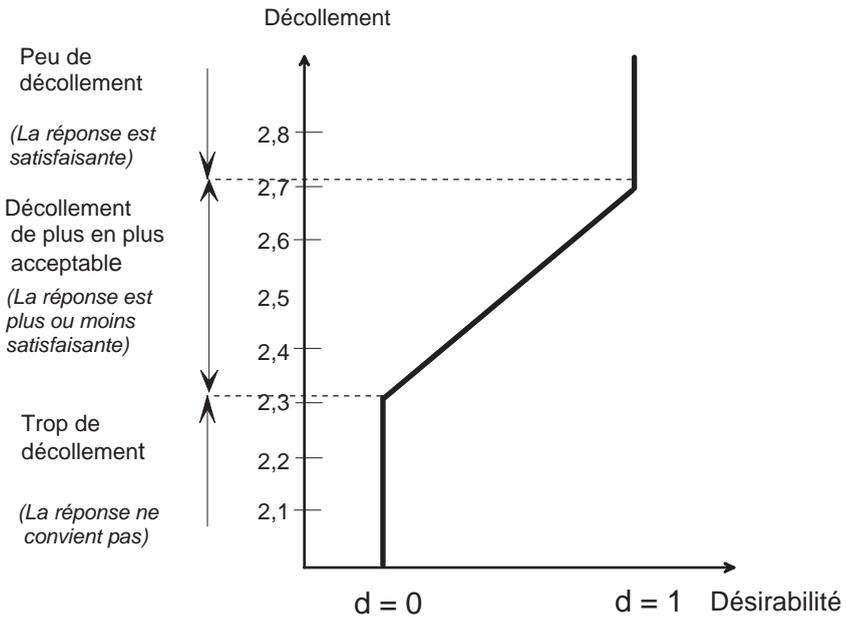


Figure 4.9 – La fonction désirabilité varie de 0 à 1 quand la réponse s’améliore progressivement.

La forme de la fonction qui permet de passer de 0 à 1 peut prendre toutes les allures possibles. Certains logiciels proposent un grand choix de telles fonctions dont certaines peuvent être compliquées pour s'adapter aux impératifs de l'étude. Dans le présent ouvrage, nous ne considérerons que les variations linéaires de la fonction désirabilité, ce qui est suffisant dans la plupart des cas.

■ Faibles valeurs souhaitées

La figure 4.9 représente une réponse dont on souhaitait les fortes valeurs. Mais on peut vouloir, au contraire, favoriser les faibles valeurs. Par exemple, s'il s'agit d'une impureté, on cherchera à en diminuer la concentration le plus possible. On adopte alors une fonction désirabilité ayant l'allure de la figure 4.10.

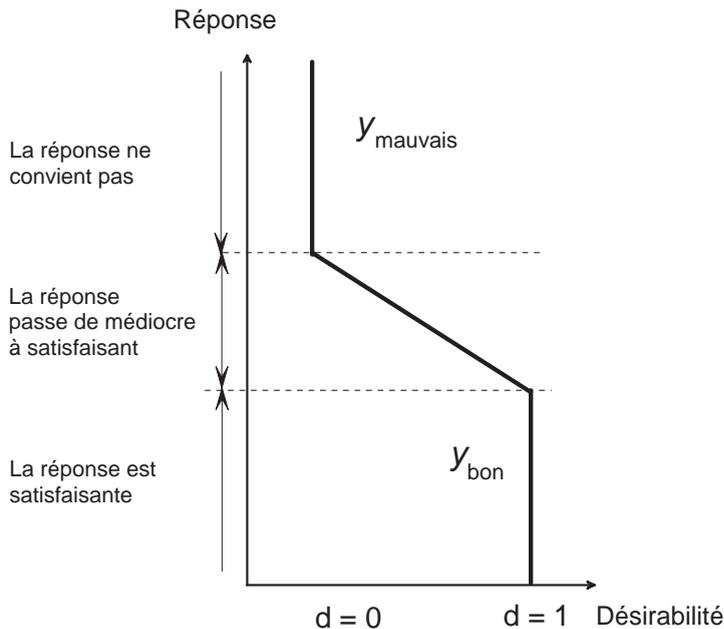


Figure 4.10 – La fonction désirabilité favorise ici les faibles valeurs de la réponse.

■ Une valeur précise souhaitée

On peut également vouloir qu'une réponse prenne une valeur précise. La figure 4.11 représente le choix d'une valeur particulière. La fonction désirabilité vaut 0 lorsqu'on est loin de la valeur souhaitée, puis elle augmente progressivement lorsqu'on en approche. Elle vaut 1 pour la valeur souhaitée, puis diminue lorsqu'on s'en éloigne pour atteindre à nouveau 0 quand on en est suffisamment loin. La montée et la descente de la fonction désirabilité peuvent ne pas avoir la même pente.

4.2.2 Plusieurs valeurs précises souhaitées

Au lieu d'une valeur précise, on peut souhaiter un ensemble de valeurs. La fonction désirabilité présente alors un plateau.

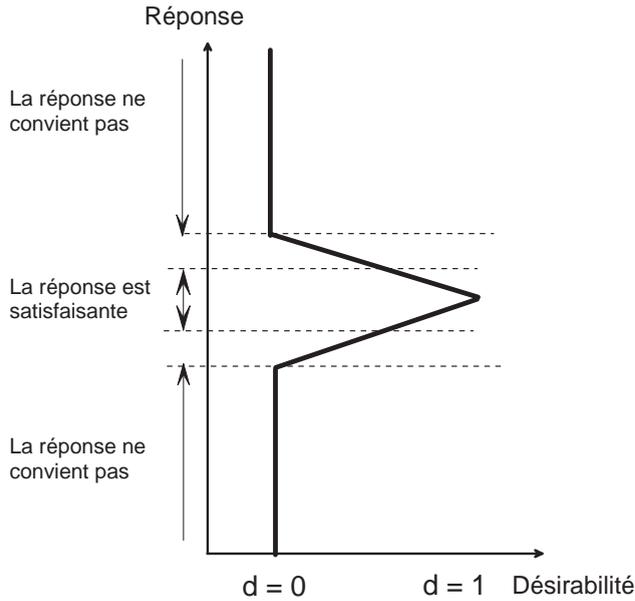


Figure 4.11 – La fonction désirabilité peut favoriser une valeur particulière.

La figure 4.12 présente un tel plateau pour lequel la vitesse de montée est différente de celle de la descente et dont les changements de pente sont amortis par des arrondis. On constate que l'on peut ajuster la forme de la fonction désirabilité à toutes les situations possibles.

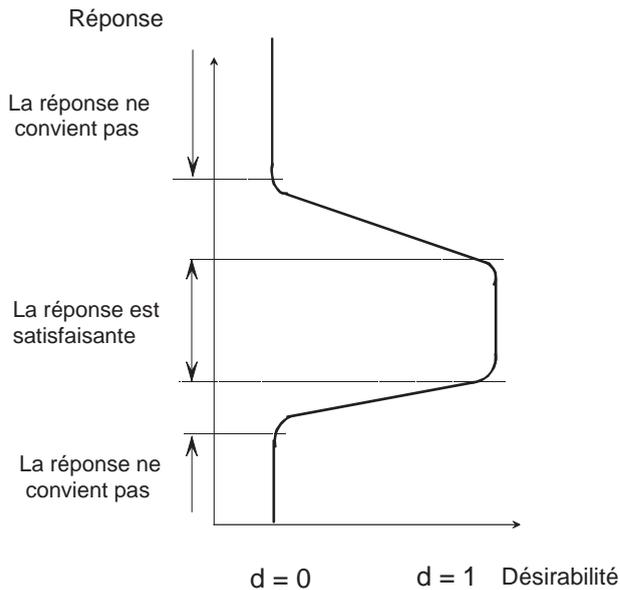


Figure 4.12 – La fonction désirabilité peut présenter un plateau.

4.2.3 La fonction désirabilité appliquée à plusieurs réponses

Nous avons examiné la fonction désirabilité pour une seule réponse. Mais l'intérêt de cette fonction est de pouvoir être utilisée pour plusieurs réponses à la fois. Il est ainsi possible de trouver le meilleur compromis lorsque des contraintes pèsent sur chacune des réponses.

La désirabilité globale D , c'est-à-dire qui prend en compte à la fois les exigences sur toutes les réponses, est la moyenne géométrique des désirabilités individuelles :

$$D = \sqrt[n]{d_1 d_2 \dots d_n}$$

On peut aboutir à plusieurs situations :

- La désirabilité globale est égale à 1 lorsque tous les objectifs sont atteints. Dans ce cas toutes les désirabilités individuelles sont égales à 1.
- La désirabilité globale est égale à 0 si l'un au moins des objectifs n'est pas atteint. Dans ce cas l'une au moins des désirabilités individuelles est égale à 0.
- La désirabilité globale est comprise entre 0 et 1 quand certains objectifs sont complètement atteints et d'autres partiellement seulement. Dans ce cas aucune désirabilité individuelle n'est égale à 0 et les désirabilités ne sont pas toutes égales à 1.

Le meilleur compromis est obtenu pour la plus forte valeur de la désirabilité globale. Les calculs d'optimisation nécessitent un logiciel programmé pour les réaliser. Nous allons regarder une application en reprenant l'exemple de la galette des rois.

4.3 Application à l'exemple 03

Supposons que les objectifs de l'étude soient un décollement supérieur à 2,5 et une fêlé supérieure à 2,8. La forme de la fonction désirabilité du décollement dépend des niveaux des 4 facteurs d'étude et elle est illustrée par la courbe de droite de la figure 4.13. Cette courbe vaut 0 jusqu'à 2, passe de 0 à 1 de 2 à 2,5 et vaut 1 au-dessus de 2,5 :

$$d_{\text{Décollement}} = f(x_1, x_2, x_3, x_4)$$

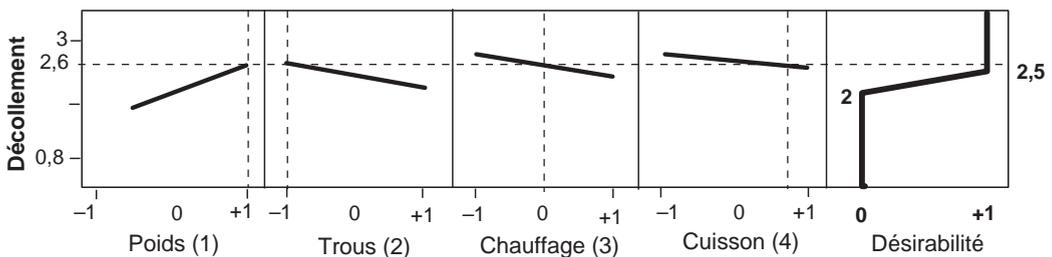


Figure 4.13 – La valeur de la fonction désirabilité du décollement varie en fonction du niveau des facteurs.

Il en est de même de la fêlé mais la fonction désirabilité est différente de celle du décollement : elle vaut 0 jusqu'à 2,5, passe de 0 à 1 de 2,5 à 2,8 et vaut 1 au-dessus de 2,8 (Figure 4.14) :

$$d_{\text{Fêlé}} = f'(x_1, x_2, x_3, x_4)$$

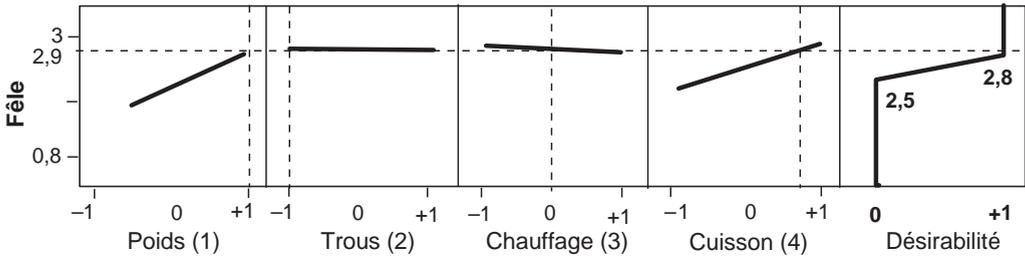


Figure 4.14 – La valeur de la fonction désirabilité de la fêle varie en fonction du niveau des facteurs.

Le logiciel calcule la valeur de la fonction désirabilité D à partir des fonctions $d_{\text{Décollement}}$ et $d_{\text{Fêle}}$. Il obtient une fonction qui dépend des niveaux des quatre facteurs et des fonctions désirabilité choisies pour chaque réponse. Le logiciel calcule la fonction désirabilité D pour chaque point du domaine d'étude. Par exemple, la figure 4.15 indique que la fonction désirabilité D vaut 0,996 pour le point (en unités codées) $x_1 = +1$, $x_2 = -1$, $x_3 = 0$, $x_4 = 0,67$. Le logiciel affiche également les coupes de la surface de réponse selon des plans définis pour des niveaux fixes des trois autres facteurs. Par exemple, pour le facteur 1 (le rapport du poids des abaisses), on a la coupe de cette surface de réponse pour les niveaux $x_2 = -1$, $x_3 = 0$, $x_4 = 0,67$ des autres facteurs. C'est en modifiant la valeur des niveaux des facteurs que l'on trouve la désirabilité maximale. Certains logiciels ont une fonction optimisation qui permet de trouver rapidement la désirabilité la plus élevée.

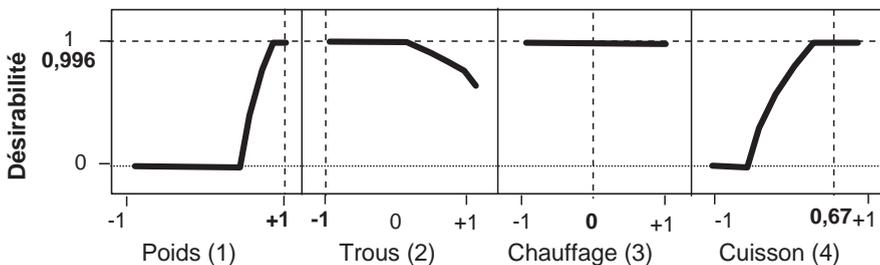


Figure 4.15 – La fonction désirabilité globale varie en fonction des niveaux des facteurs.

On obtient ainsi les niveaux des facteurs donnant la désirabilité la plus forte. Comme on peut obtenir une désirabilité égale à 1 pour différents réglages des niveaux des facteurs, l'optimiseur ne donne pas toujours la même solution. Il en donne une parmi toutes les solutions possibles. Il donne celle qui est la plus proche de celle que vous avez trouvée par tâtonnements lorsque vous avez déplacé le curseur sur les diagrammes.

4.4 Conclusion de l'étude

On diminue le décollement et la fêle en réglant :

- le rapport du poids des abaisses (poids de l'abaisse supérieure/poids de l'abaisse inférieure) à 1,1 ;

- le nombre de trous au niveau « Peu » ;
- la température du four au niveau « Forte au début ».

On favorise la diminution du décollement au détriment de la fêle en réduisant la durée de cuisson à 29 min. Inversement, on favorise la fêle au détriment du décollement en augmentant la durée de cuisson à 31 min.

Des voies d'amélioration sont possibles si l'on sort du domaine d'étude défini au début de l'étude. En effet, on constate que l'on pourrait atteindre les objectifs en augmentant le rapport du poids des abaisses au-delà de 1,1, en diminuant le nombre de trous et en augmentant la température du four en début de chauffe. Il s'agit là d'une *extrapolation maîtrisée* : on utilise le modèle mathématique établi dans le domaine d'étude en dehors de ce domaine. On fait donc l'hypothèse que, dans le voisinage du domaine d'étude, les tendances indiquées par le modèle mathématique sont encore acceptables. Il faut bien sûr, comme on le fait pour toutes les hypothèses, vérifier que les prévisions extrapolées sont confirmées par l'expérience. Dans le cas de la galette des rois, des essais complémentaires ont été entrepris pour vérifier que l'on atteignait bien la note 3 pour les deux réponses (le décollement et la fêle) lorsqu'on sortait du domaine initialement choisi. Les essais ont été essentiellement orientés vers une légère augmentation du rapport du poids des abaisses et vers une augmentation de la température en début de four. L'objectif a été atteint pour :

- le rapport du poids des abaisses (poids de l'abaisse supérieure/poids de l'abaisse inférieure) à 1,2 ;
- le nombre de trous au niveau « Peu » (sans changement) ;
- la température du four au niveau « Un peu plus forte que précédemment ».

La pratique montre que l'extrapolation maîtrisée est un moyen efficace pour orienter une étude vers la solution souhaitée. Mais, il est indispensable, pour éviter de mauvaises surprises, de vérifier que les prévisions extrapolées sont bien confirmées par l'expérience.

5 • NOTIONS DE STATISTIQUE APPLIQUÉES AUX PLANS D'EXPÉRIENCES

Les plans factoriels complets que nous avons vus permettent d'évaluer l'influence des facteurs et de leurs interactions par l'intermédiaire des coefficients du modèle postulé. Nous avons estimé que les coefficients ayant la plus forte valeur étaient les plus importants. Nous avons jugé l'importance des coefficients en les comparant entre eux. Mais nous n'avons aucun étalon pour comparer tous les coefficients et indiquer ceux qui devaient être conservés et ceux qui pouvaient être éliminés du modèle mathématique final. Ce qui nous manquait c'était une évaluation de l'incertitude attachée aux coefficients. Il y a deux sources d'erreurs sur la détermination des coefficients d'un modèle. La première est l'erreur commise sur chaque réponse, c'est l'erreur expérimentale. La seconde est le modèle mathématique lui-même qui n'est qu'une représentation plus ou moins bonne du phénomène étudié, c'est le manque d'ajustement. L'incertitude est constituée de ces erreurs.

Quand on réalise plusieurs fois la même expérience dans les mêmes conditions opératoires et que l'on mesure les réponses, on trouve des valeurs qui sont proches les unes des autres mais qui sont différentes de l'une à l'autre. Il y a une dispersion des mesures. Cette dispersion traduit le fait que les réponses sont des grandeurs aléatoires. Il y a une erreur expérimentale attachée à chaque détermination d'une réponse. Cette erreur expérimentale se transmet aux coefficients.

Il y a également une différence entre le modèle postulé par l'expérimentateur et le modèle réel. Cette différence est le manque d'ajustement entre les deux modèles. Cette erreur se transmet également aux coefficients du modèle. Ce sont ces deux types d'erreur, l'erreur expérimentale et l'erreur d'ajustement, qui vont servir pour évaluer l'importance des coefficients. Par exemple, si l'erreur d'appréciation sur un coefficient est de 5 et si ce coefficient vaut 4, il sera considéré comme négligeable et non influent. En revanche, un coefficient de valeur 30 sera considéré comme influent. Il est donc important de connaître l'erreur que l'on commet dans l'estimation des coefficients. Pour mener à bien cette étude, il faut connaître les bases statistiques qui sont utilisées pour faire les calculs et appliquer les raisonnements corrects. Les logiciels nous soulagent de toute la partie calcul, mais il faut être capable d'interpréter les grandeurs statistiques qu'ils nous fournissent.

L'exemple des outils de coupe jetables servira à illustrer les concepts théoriques qui sont développés dans ce chapitre. Nous introduirons d'abord les deux grandeurs statistiques les plus importantes, la moyenne et l'écart-type. Puis, nous verrons les notions de population, d'échantillon et de distribution. Nous insisterons sur le rôle

prépondérant de l'écart-type qui sert d'étalon pour évaluer l'importance d'une grandeur aléatoire. Nous verrons comment l'écart-type permet de calculer la *p-value* qui mesure la probabilité qu'un événement soit rare. L'objectif premier de ce chapitre est de préciser les notions qui permettent d'évaluer et d'améliorer l'aptitude du modèle postulé à représenter les résultats expérimentaux. En un mot ces notions nous aideront à trouver le modèle qui représente le mieux possible les données expérimentales et qui devrait nous permettre de faire des prévisions de qualité.

5.1 Exemple 04 : durée de vie des outils de coupe jetables

5.1.1 Préparation du plan d'expériences

■ Description de l'étude

L'usinage des métaux nécessite des outils de coupe dont la matière et les profils sont particulièrement étudiés. Les bons outils doivent couper parfaitement, ne pas mettre en danger la sécurité des ouvriers, ne pas s'user trop vite et assurer la précision des objets fabriqués. L'affûtage de tels outils pose de nombreux problèmes et un fabricant a décidé d'étudier des outils jetables. Pour assurer leur rentabilité ces outils doivent avoir une durée de vie suffisante.

La présente étude porte sur des outils spéciaux à usage intensif dont on cherche à maximiser la durée de vie.

■ Objectif de l'étude

L'objectif de l'étude est de trouver les conditions d'emploi qui assurent une durée de vie supérieure à 20 heures aux outils de coupe.

■ Choix de la réponse permettant d'atteindre l'objectif

La réponse choisie est la durée de vie des outils.

■ Facteurs et domaine d'étude

Les facteurs retenus sont :

- le débit de l'huile de coupe,
- la vitesse de coupe de l'outil,
- la profondeur de coupe de l'outil,
- la vitesse d'avance de l'outil.

Soit quatre facteurs dont les domaines sont indiqués dans le tableau 5.1.

Tableau 5.1 – Facteurs retenus et domaine d'étude.

Facteurs	Niveau -1	Niveau 0	Niveau +1
Débit (1)	650	725	800
Vitesse de coupe (2)	10	18	26
Profondeur (3)	0,05	0,125	0,20
Vitesse d'avance (4)	0,5	0,75	1

■ Choix du plan d'expériences

L'expérimentateur décide d'exécuter un plan complet 2^4 , soit 16 essais. Mais il désire avoir une estimation de l'erreur expérimentale. Pour ce faire, il ajoute quatre essais situés au centre du domaine d'étude. Il va donc effectuer 20 essais (Tableau 5.2).

5.1.2 Expérimentation

Tableau 5.2 – Essais et réponses du plan d'expériences « Outils de coupe jetables » (unités codées).

Essai n°	Débit (1)	Vitesse de coupe (2)	Profondeur (3)	Vitesse d'avance (4)	Durée de vie (h)
1	-1	-1	-1	-1	26,1
2	+1	-1	-1	-1	22,2
3	-1	+1	-1	-1	10,1
4	+1	+1	-1	-1	12,2
5	-1	-1	+1	-1	14,2
6	+1	-1	+1	-1	12,7
7	-1	+1	+1	-1	5,9
8	+1	+1	+1	-1	5,6
9	-1	-1	-1	+1	23
10	+1	-1	-1	+1	20,1
11	-1	+1	-1	+1	2,4
12	+1	+1	-1	+1	3,7
13	-1	-1	+1	+1	11
14	+1	-1	+1	+1	13,4
15	-1	+1	+1	+1	0,5
16	+1	+1	+1	+1	1,7
17	0	0	0	0	11,1
18	0	0	0	0	12,6
19	0	0	0	0	10,4
20	0	0	0	0	11,9
-1	650	10	0,05	0,5	
0	725	18	0,125	0,75	
+1	800	26	0,2	1	

Les points au centre sont regroupés à la fin du plan pour la présentation des essais. Mais l'ordre d'exécution des essais d'un plan ne suit pas forcément l'ordre de présentation. On pourra par exemple faire un point au centre au début de l'étude, un autre après l'essai n° 6, un troisième après l'essai n° 12 et le dernier à la fin.

5.1.3 Calculs et interprétation

On adopte un modèle avec les facteurs principaux et les interactions entre deux facteurs. Les interactions entre trois et quatre facteurs ne rentrent pas dans ce modèle. Pour faire ces calculs on ne tient compte que des 16 essais du plan factoriel. On ne prend pas en compte les 4 points centraux. En effet, ces points sont des points de contrôle et non des points de calcul. Ils serviront à calculer une estimation de l'erreur expérimentale.

Le logiciel calcule le coefficient constant, les effets principaux et les six interactions d'ordre 2 (Tableau 5.3).

Tableau 5.3 – Coefficients du modèle (unités codées).

Effet	Valeur	Écart-type	t de Student	p-value
Constante	11,55	0,3609	32	< 0,0001
Débit (1)	-0,10	0,3609	-0,28	0,7928
Vitesse de coupe (2)	-6,29	0,3609	-17,42	< 0,0001
Profondeur (3)	-3,42	0,3609	-9,49	0,0002
Vitesse d'avance (4)	-2,07	0,3609	-5,75	0,0022
12	0,64	0,3609	1,77	0,1376
13	0,32	0,3609	0,90	0,4091
14	0,35	0,3609	0,97	0,3767
23	1,59	0,3609	4,40	0,0070
24	-1,11	0,3609	-3,08	0,0274
34	0,60	0,3609	1,66	0,1573

Pour bien interpréter ce tableau, il faut comprendre la signification des informations fournies, en particulier l'écart-type, le t de Student et la p -value. Cela nous oblige à faire un long détour pour introduire les principales notions de statistiques utiles aux plans d'expériences. Nous verrons comment on utilise les principales propriétés de la distribution normale pour estimer les erreurs et les probabilités. Ces outils statistiques permettent d'obtenir une bonne méthode d'évaluation de l'influence des facteurs. Enfin nous verrons comment l'analyse de la variance

donne accès à des grandeurs statistiques utiles pour bien évaluer les modèles mathématiques utilisés. La dernière partie de cette grande parenthèse est une application qui permet de voir comment sont utilisées les notions de statistique introduites.

5.2 Notions de statistique appliquées aux plans d'expériences

On définira d'abord l'erreur expérimentale, puis les notions de population et d'échantillon.

5.2.1 Erreur expérimentale

Les quatre points au centre ont des valeurs différentes. Au lieu de donner la liste des quatre valeurs, on peut essayer de la résumer en indiquant la valeur centrale et la dispersion autour de cette valeur centrale. En général, on prend la moyenne arithmétique comme valeur centrale et l'écart-type comme mesure de la dispersion.

■ Moyenne

Par définition, la moyenne arithmétique d'un ensemble de valeurs est la somme de toutes les valeurs divisées par le nombre de valeurs. Ici, la moyenne arithmétique est égale à :

$$\bar{y} = \frac{1}{4}(11,1 + 12,6 + 10,4 + 11,9) = 11,5$$

■ Écart-type

La définition de l'écart-type est un peu moins simple que celle de la moyenne. Nous allons décrire son calcul pas à pas :

1. On commence par calculer les écarts à la moyenne, c'est-à-dire la différence entre chaque valeur et la moyenne :

$$11,1 - 11,5 = -0,4$$

$$12,6 - 11,5 = +1,1$$

$$10,4 - 11,5 = -1,1$$

$$11,9 - 11,5 = 0,4$$

Notez que la somme de ces écarts à la moyenne est égale à zéro. On démontre d'ailleurs que c'est toujours le cas :

$$\sum_{i=1}^{i=n} (y_i - \bar{y}) = \sum_{i=1}^{i=n} y_i - \sum_{i=1}^{i=n} \bar{y} = \sum_{i=1}^{i=n} y_i - n\bar{y} = n\bar{y} - n\bar{y} = 0$$

On ne peut donc pas prendre la somme des écarts comme mesure de la dispersion. C'est pourquoi on fait disparaître le signe négatif en prenant les carrés des écarts.

2. Ces écarts à la moyenne sont donc élevés au carré et additionnés. On obtient ainsi la somme des carrés des écarts à la moyenne :

$$(-0,4)^2 + (1,1)^2 + (-1,1)^2 + (0,4)^2 = +0,16 + 1,21 + 1,21 + 0,16 = +2,74$$

3. Cette somme est divisée par le nombre de valeurs moins 1 :

$$\frac{2,74}{4 - 1} = 0,9133$$

Cette quantité porte le nom de *variance*. C'est une grandeur fondamentale de la science statistique. On la retrouve partout et il en sera fait un grand usage.

4. Enfin l'*écart-type* est obtenu en prenant la racine carrée de la variance :

$$\text{écart-type} = \sqrt{0,9133} = 0,9557 \cong 0,96$$

Pourquoi prend-on la racine carrée de la variance ? Simplement pour exprimer la dispersion dans la même unité que les données d'origine et que la moyenne. Les comparaisons sont ainsi facilitées. En effet, quand on a élevé au carré les écarts à la moyenne, on a aussi élevé au carré l'unité de mesure. La variance est donc exprimée avec une unité qui a pour dimension le carré de l'unité d'origine.

Heureusement les ordinateurs sont là pour faire ces calculs et ils indiquent directement la valeur de l'écart-type des réponses.

Que signifient ces deux grandeurs, moyenne et écart-type ? Et comment allons-nous nous en servir ?

Si l'on a une série de réponses effectuées au même point expérimental, on peut calculer la moyenne et l'écart-type de la série. Mais peut-on dire d'avance, avec une forte probabilité de ne pas se tromper, quelle serait la valeur d'une nouvelle mesure effectuée au même point expérimental ? Il est fort probable qu'elle ne serait pas très différente de la moyenne. Si au lieu d'indiquer un chiffre, on indiquait une fourchette, on aurait moins de chances de se tromper. Si l'on dit : « la valeur sera probablement comprise entre telle et telle valeur », on augmente les chances de prévision par rapport à la moyenne. Si on prend une fourchette plus large, on aura de grandes chances de faire une bonne prévision. Les statisticiens ont calculé la probabilité que l'on avait de ne pas se tromper ! Et cela grâce à l'écart-type.

Pour approfondir ces importantes questions, nous allons voir successivement les notions de *population*, d'*échantillon*, de *distribution*, de *degrés de liberté* et de *transmission des erreurs*.

5.2.2 Population

Supposons que nous fassions plusieurs mesures dans les mêmes conditions expérimentales. Nous obtenons une suite de valeurs qui sont proches les unes des autres mais qui sont légèrement différentes. Imaginons que nous poursuivions les mesures jusqu'à en obtenir une infinité. L'ensemble de toutes ces valeurs forme une population de grandeurs aléatoires caractérisée par :

- la valeur centrale des mesures, appelée *moyenne de la population* et notée μ ;
- l'écart-type de la population, noté $\sigma_{\text{population}}$
- la distribution.

5.2.3 Échantillon

D'un point de vue statistique, obtenir un échantillon de n valeurs, c'est tirer au hasard n valeurs dans la population de toutes les valeurs possibles. La valeur de n

est petite vis-à-vis du nombre d'individus de la population. Un échantillon de n valeurs est caractérisé par :

- la valeur moyenne des mesures ; soit un échantillon de n réponses, y_i ; la moyenne \bar{y} est calculée grâce à la relation :

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{i=n} y_i$$

- l'écart-type de l'échantillon, noté $\sigma_{\text{échantillon}}$;
- l'histogramme des valeurs.

Un histogramme est une représentation graphique de l'ensemble de toutes les valeurs de l'échantillon. Pour construire l'histogramme de la figure 5.1, nous avons supposé que l'échantillon contenait 100 valeurs comprises entre 9 et 14. On définit des classes de largeur 0,5 unité. La première classe contient 1 valeur comprise entre 9 et 9,50. La deuxième contient 5 valeurs comprises entre 9,50 et 10. La troisième contient 8 valeurs comprises entre 10 et 10,50. La quatrième contient 14 valeurs comprises entre 10,50 et 11. La cinquième contient 20 valeurs comprises entre 11 et 11,50. La sixième classe contient 22 valeurs comprises entre 11,50 et 12. Les classes suivantes contiennent 15, 9, 5 valeurs et la dernière contient 1 valeur.

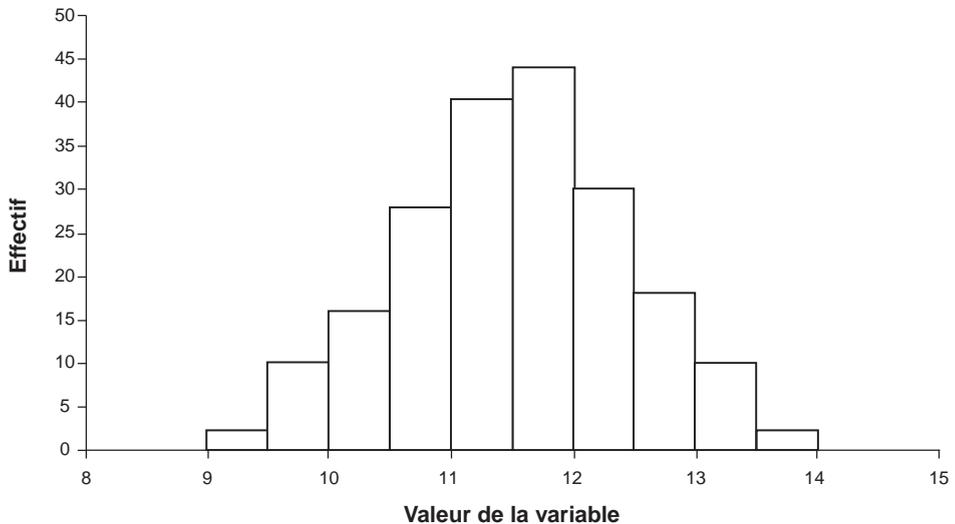


Figure 5.1 – Histogramme des 100 mesures de l'échantillon.

5.2.4 De l'échantillon à la population

Il est très rare que l'on puisse connaître directement les caractéristiques d'une population. Le moyen le plus souvent utilisé est d'estimer ces caractéristiques à partir d'un échantillon. Les statisticiens se sont penchés sur ce problème et voici leurs conclusions :

- La meilleure estimation de la moyenne de la population, c'est la moyenne de l'échantillon :

Meilleure estimation de $\mu = \bar{y}$ de l'échantillon

- La meilleure estimation de l'écart-type de la population est notée par la lettre s . Elle est donnée par la relation :

$$s = \text{Meilleure estimation de } \sigma_{\text{population}} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{i=n} (y_i - \bar{y})^2}$$

Il faut souligner que le dénominateur est $n - 1$ et non n .

■ Degrés de liberté

Soit n réponses mesurées indépendamment les unes des autres. Il n'existe pas de relation mathématique entre elles. Les n écarts à la moyenne correspondants ne sont pas indépendants. En effet, il existe une relation mathématique entre ces écarts. Quand on en connaît $n - 1$, on peut calculer le dernier avec la relation mathématique. Par exemple, reprenons les quatre écarts à la moyenne de l'exemple (voir l'écart-type). Les trois premiers écarts sont :

$$-0,4 \qquad +1,1 \qquad -1,1$$

Le quatrième écart s'obtient facilement puisque la somme des écarts est toujours égale à 0 :

$$\begin{aligned} \text{quatrième écart} - 0,4 + 1,1 - 1,1 &= 0 \\ \text{quatrième écart} &= 0,4 \end{aligned}$$

Il n'y a donc que $n - 1$ écarts indépendants. On dit que la série des n écarts à la moyenne possède $n - 1$ *degrés de liberté* (ou ddl). Le nombre de degrés de liberté est important car il intervient dans de nombreuses formules de statistique.

■ Distribution

Au fur et à mesure que l'on ajoute des valeurs à un échantillon, on obtient un histogramme de plus en plus régulier et, pour un nombre infini de valeurs, c'est-à-dire pour une population, on a une répartition qui s'appelle une *distribution*. Il est possible de décrire la distribution d'une population par une relation mathématique. Par exemple, la distribution la plus souvent rencontrée dans les plans d'expériences est la distribution normale ou distribution de Laplace-Gauss. Cette distribution est illustrée par une courbe ayant l'aspect d'une cloche et elle est définie par la relation :

$$f(y) = \frac{1}{\sigma_{\text{population}} \sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma_{\text{population}}^2}}$$

Cette distribution est parfaitement définie par la moyenne de la population μ et par l'écart-type $\sigma_{\text{population}}$. On peut ramener toutes les distributions de Laplace-Gauss à la distribution normale ayant une moyenne nulle et un écart-type unité (Figure 5.2).

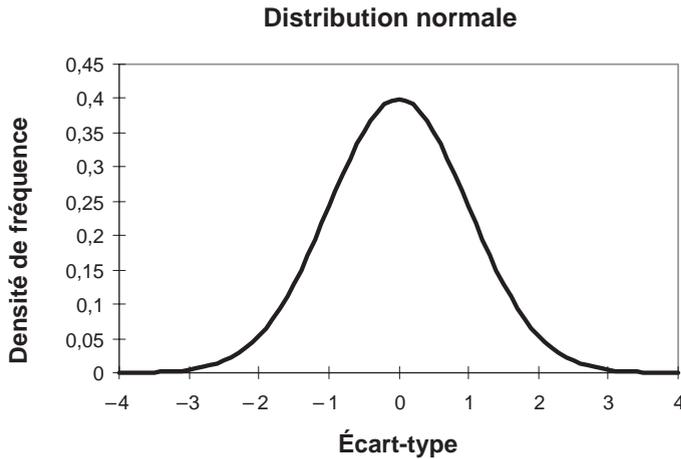


Figure 5.2 – Représentation de la distribution normale
($\mu = 0$; $\sigma_{population} = 1$).

5.2.5 Principales propriétés de la distribution normale

La lecture de la courbe de Laplace-Gauss consiste à regarder la surface qui est sous la courbe en cloche. La surface totale représente toutes les valeurs possibles de la population. La probabilité qu'une valeur soit supérieure à 1,2 (ou toute autre valeur donnée) est représentée par la surface comprise entre 1,2 et l'infini (Figure 5.3).

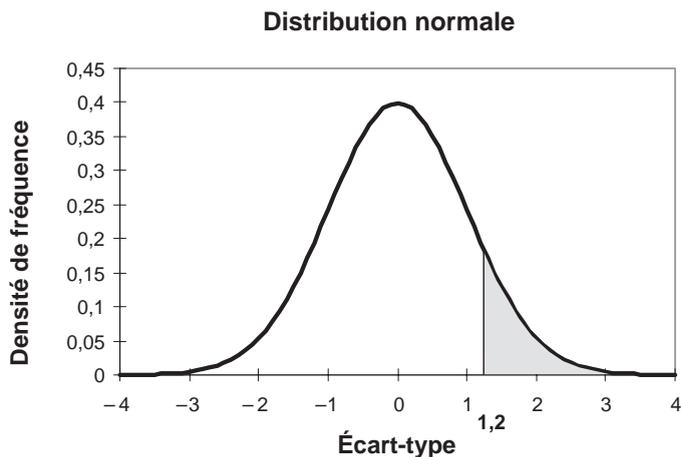


Figure 5.3 – La probabilité de tirer une valeur supérieure à 1,2
est proportionnelle à la surface grisée.

La probabilité qu'une valeur soit inférieure à 0,8 (ou toute autre valeur donnée) est représentée par la surface comprise entre moins l'infini et 0,8 (Figure 5.4).

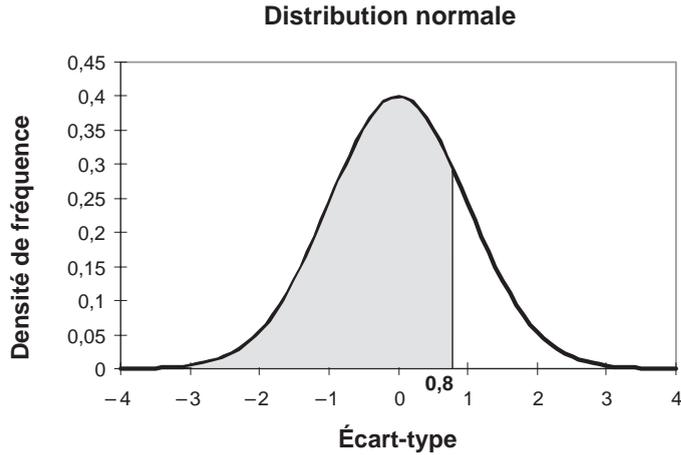


Figure 5.4 – La probabilité de tirer une valeur inférieure à 0,8 est proportionnelle à la surface grisée.

Enfin la probabilité qu'une valeur soit comprise entre $-0,5$ et $+1,5$ (ou toute autre valeur donnée) est représentée par la surface comprise entre ces deux valeurs (Figure 5.5).

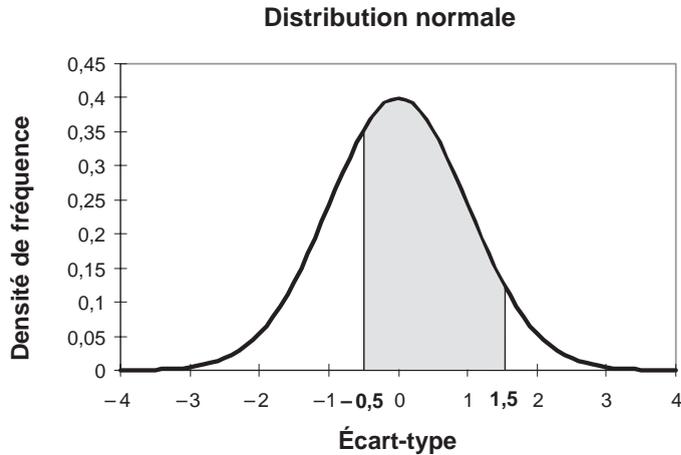


Figure 5.5 – La probabilité de tirer une valeur comprise entre $-0,5$ et $+1,5$ est proportionnelle à la surface grisée.

Par exemple, il a été démontré que 68,26 % des valeurs de la population sont comprises entre moins un écart-type et plus un écart-type autour de la moyenne de la population (Figure 5.6). On en déduit que si l'on fait un tirage au hasard dans une population de distribution normale, on a environ 68 % de chances de tirer une valeur comprise entre $\mu \pm \sigma_{population}$ et 32 % de chances de tirer une valeur extérieure à cet intervalle.

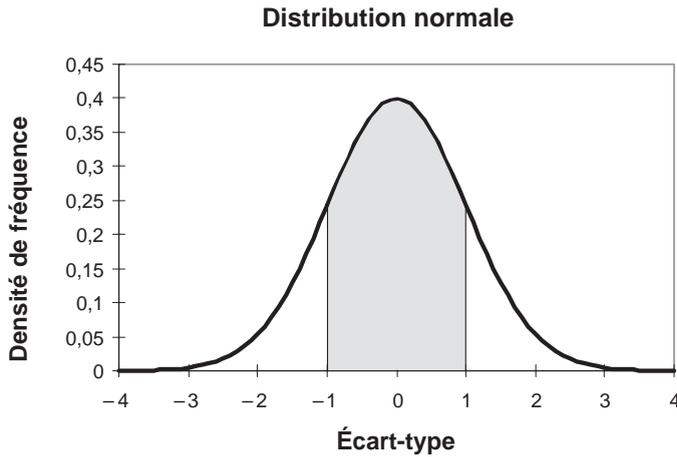


Figure 5.6 – Dans l'intervalle $\mu \pm \sigma_{\text{population}}$, il y a 68,26 % de la population.

On a montré de même que 95 % des valeurs de la population sont comprises entre moins 1,96 écart-type et plus 1,96 écart-type autour de la moyenne de la population (Figure 5.7). Un tirage au hasard dans une population de distribution normale aura une valeur comprise entre $\mu \pm 1,96\sigma_{\text{population}}$ dans 95 % des cas. Dans 5 % des cas, on tirera une valeur extérieure à cet intervalle.

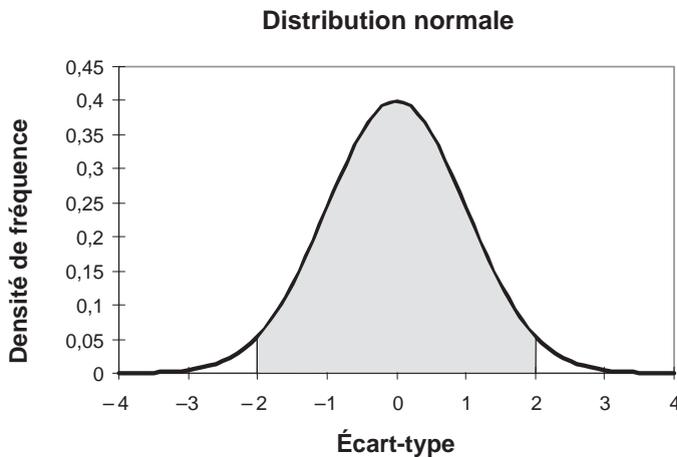


Figure 5.7 – Dans l'intervalle $\mu \pm 1,96\sigma_{\text{population}}$, il y a 95 % de la population.

Enfin dans l'intervalle $\mu \pm 3\sigma_{\text{population}}$, il y a 99,74 % des valeurs de la population (Figure 5.8). Il y a environ 2,5 chances sur 1 000 de tirer une valeur qui diffère de 3 écarts-types de la moyenne de la population.

Dans tous les cas, c'est l'écart-type de la population qui est pris comme référence. Cette statistique est donc très importante puisque c'est grâce à elle que l'on peut estimer la dispersion des mesures. On prend l'estimation de l'écart-type de la population des réponses comme mesure de l'erreur expérimentale.

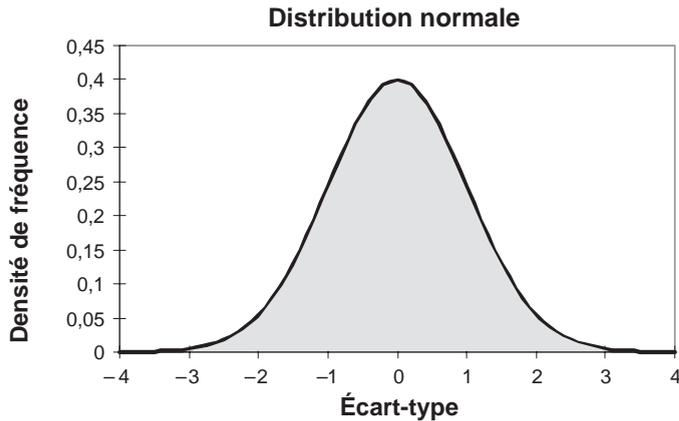


Figure 5.8 – Dans l'intervalle $\mu \pm 3\sigma_{population}$ il y a 99,74 % de la population.

5.2.6 Transmission des erreurs

D'un point de vue statistique, déterminer une réponse, c'est faire un tirage au hasard d'une valeur dans la population de ces réponses. Cette population possède une distribution, une moyenne et un écart-type. Dans la majorité des cas, la distribution des réponses est une distribution de Laplace-Gauss et l'écart-type de cette population est une mesure de l'erreur expérimentale. Si l'on tire plusieurs réponses au hasard, elles ont de grandes chances d'être réparties autour de la moyenne.

Si on a réalisé plusieurs mesures de la réponse au point central, on peut calculer la moyenne de ces réponses :

$$\bar{y} = \frac{1}{4}[y_1 + y_2 + y_3 + y_4]$$

Comment les erreurs commises sur les réponses y_i se répercutent-elles sur la moyenne ? Les statisticiens ont étudié ce problème et se sont aperçus que les erreurs ne s'additionnaient pas. L'erreur commise sur la moyenne n'est pas quatre fois l'erreur commise sur une réponse. Pour résoudre ce problème, il faut faire intervenir une nouvelle grandeur : la *variance*. Par définition, la variance d'une réponse $V(y_i)$ est le carré de son écart-type, c'est-à-dire que si l'écart-type de la réponse y_1 est σ_y , la variance de cette réponse est σ_y^2 . La relation donnant une estimation de la variance s_y^2 de la population des réponses est :

$$s_y^2 = V(y_i) = \frac{1}{n-1} \sum_{i=1}^{i=n} (y_i - \bar{y})^2$$

On constate que la variance est égale à la somme des carrés des écarts à la moyenne divisée par le nombre de degrés de liberté des écarts à la moyenne. D'une manière générale, la variance d'une série de données est obtenue de la même manière : on divise la somme des carrés des écarts à la moyenne par le nombre de degrés de liberté de la série de données. Cette règle sera largement utilisée dans le paragraphe sur l'analyse de la variance où plusieurs variances ayant des degrés de liberté différents seront calculées.

Ce sont les variances qui s'additionnent et elles s'additionnent quel que soit le signe (positif ou négatif) qui est devant la réponse. La variance d'une somme est donc égale à :

$$V(+y_1 + y_2 + y_3 + y_4) = V(y_1) + V(y_2) + V(y_3) + V(y_4)$$

La variance d'une différence est égale à :

$$V(-y_1 + y_2 - y_3 + y_4) = V(y_1) + V(y_2) + V(y_3) + V(y_4)$$

Dans la formule de la moyenne, il y a le coefficient $1/4$. Quand on passe à la variance, ce coefficient est élevé au carré. La variance de la moyenne est égale à :

$$V(\bar{y}) = \left(\frac{1}{4}\right)^2 V(y_1 + y_2 + y_3 + y_4) = \left(\frac{1}{4}\right)^2 [V(y_1) + V(y_2) + V(y_3) + V(y_4)]$$

Si les variances des réponses sont égales entre elles, on dit qu'il y a *homoscédasticité* et la relation précédente devient :

$$V(\bar{y}) = \left(\frac{1}{4}\right)^2 [V(y_1) + V(y_2) + V(y_3) + V(y_4)] = \left(\frac{4}{16}\right) V(y_i)$$

$$V(\bar{y}) = \frac{1}{4} V(y_i)$$

Prenons les racines carrées des deux membres de cette relation et remplaçons par n le nombre de réponses intervenant dans la formule de la moyenne :

$$\sqrt{V(\bar{y})} = \sqrt{\frac{V(y_i)}{n}}$$

soit :

$$\sigma_{\bar{y}} = \frac{\sigma_{y_i}}{\sqrt{n}} \quad \{5.1\}$$

L'écart-type de la moyenne est égal à l'écart-type des réponses individuelles divisé par la racine carrée du nombre des réponses individuelles qui interviennent dans le calcul de la moyenne. Cette relation indique que plus n est grand, meilleure est la précision sur la moyenne.

La dispersion de la moyenne est plus faible que la dispersion des réponses. On a donc une meilleure précision sur l'estimation de la moyenne que sur les réponses elles-mêmes.

5.2.7 Intervalle de confiance des réponses

Nous venons de voir l'importance de l'écart-type de la population des réponses qui est considéré comme une évaluation de l'erreur expérimentale ou erreur pure. L'estimation de l'écart-type s d'une population de réponses à partir d'un échantillon est obtenue à l'aide de la relation :

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{i=n} (y_i - \bar{y})^2}$$

Mais la qualité de cette estimation dépend du nombre n de valeurs prises en compte pour calculer l'écart-type. Si l'échantillon contient beaucoup de mesures, l'écart-type calculé sera de bonne qualité, c'est-à-dire proche de l'écart-type de la population. Si l'échantillon contient peu de mesures, l'écart-type calculé sera de moins bonne qualité et il faudra prendre des précautions supplémentaires pour l'utiliser. L'intervalle de confiance I_c dépend :

- de la probabilité choisie par l'utilisateur ; cette probabilité est souvent de 95 ou 99 % ;
- de la qualité de l'écart-type calculé.

On a :

$$I_c = k \sigma_{\text{échantillon}}$$

Le tableau 5.4 indique les valeurs de k en fonction de la probabilité choisie par l'utilisateur (pourcentage de cas où l'hypothèse est correcte) et de la qualité de l'écart-type calculé (nombre de mesures ayant servi à calculer s).

Si l'échantillon possède un très grand nombre de valeurs indépendantes, la qualité de l'écart-type est bonne. On observe cette situation dans les usines ou les laboratoires de contrôle qui procèdent journellement aux mêmes mesures sur les mêmes produits. Choisissons une probabilité de 95 %. Dans ce cas, si l'on effectue une mesure y_i , on pourra dire que l'on a 95 % de chances que l'intervalle $y_i \pm 1,96 \sigma_{\text{échantillon}}$ contienne la moyenne de la population.

Dans le cas des plans d'expériences, il est rare que n soit élevé. On a, le plus souvent, réalisé quelques mesures de la réponse au même point d'expériences, en général au centre du domaine expérimental. Il s'agit de répétitions, c'est-à-dire de mesures réalisées dans les mêmes conditions expérimentales. Supposons que l'écart-type ait été obtenu avec 5 mesures de répétition et que l'on souhaite une probabilité de 95 %. Dans ce cas, si l'on effectue une mesure y_i , on aura 95 % de chances que l'intervalle $y_i \pm 2,78s$ contienne la moyenne de la population. Avec 10 mesures de répétition pour le calcul de l'écart-type, on pourra dire que l'on a 95 % de chances que l'intervalle $y_i \pm 2,26 s$ contienne la moyenne de la population.

Tableau 5.4 – Nombre d'écart-types à prendre compte pour déterminer la probabilité que la moyenne de la population soit dans l'intervalle de confiance.

		Nombre de mesures ayant servi à calculer s						
		2	3	4	5	10	20	∞
Pourcentage de cas où l'hypothèse est correcte	70 %	1,96	1,38	1,25	1,19	1,10	1,06	1,03
	90 %	6,31	2,92	2,35	2,13	1,83	1,73	1,64
	95 %	12,71	4,3	3,18	2,78	2,26	2,06	1,96
	99 %	63,66	9,92	5,84	4,60	3,25	2,86	2,58

5.3 Facteurs et interactions influents

L'influence des facteurs et de leurs interactions est évaluée par l'intermédiaire des coefficients du modèle postulé. Il nous faut donc trouver une valeur étalon pour juger de l'importance d'un coefficient. Cet étalon sera l'écart-type du coefficient. On comparera la valeur du coefficient à la valeur de son écart-type.

1. Si le coefficient est nettement plus grand que son écart-type, on estimera que le coefficient est influent :

$$a_1 \gg \sigma_{a_1}$$

2. Si le coefficient est nettement plus petit que son écart-type, on estimera que l'on peut éliminer ce coefficient :

$$a_1 \ll \sigma_{a_1}$$

3. Si le coefficient est du même ordre de grandeur que son écart-type, il faudra faire appel à ses connaissances du phénomène et aux enjeux de l'étude pour prendre une décision :

$$a_1 \cong \sigma_{a_1}$$

Pour évaluer l'importance d'un coefficient, la théorie des statistiques compare ce coefficient à son écart-type en faisant le rapport a_1/σ_{a_1} . Ce rapport est appelé le *t de Student*. À partir du *t* de Student, on peut évaluer la probabilité que le coefficient soit nul ou, autrement dit, sans importance ou peu significatif. Cette probabilité est la *p-value*. Si la *p-value* est proche de 0 (probabilité nulle), le coefficient est influent car il n'est pas nul. Si la *p-value* est proche de 1, le coefficient est proche de 0 et donc négligeable. Si la *p-value* possède une valeur intermédiaire, le coefficient est peut-être légèrement significatif ou non significatif.

Savoir si un coefficient est négligeable ou non impose d'avoir une évaluation de l'écart-type de la population de ce coefficient. Nous savons que les coefficients dépendent des réponses, du modèle postulé et de la méthode de calcul. De même, la variance des coefficients dépend des réponses, du modèle postulé et de la méthode de calcul. Les statisticiens ont établi les formules donnant la variance des coefficients $V(a_i) = \sigma_{a_i}^2$ en fonction de la variance des écarts $V(e)$, somme de la variance de l'erreur expérimentale et de la variance de l'erreur commise sur le modèle postulé :

$$V(a_i) = K V(e)$$

où la constante *K* dépend du modèle postulé et de la matrice d'expériences. Dans le cas le plus général *K* est très long à calculer et seuls les logiciels spécialisés possèdent les algorithmes pour faire ce calcul. Dans le cas des plans factoriels dont les niveaux -1 et $+1$ sont parfaitement respectés et lorsque le modèle postulé est un polynôme, c'est-à-dire dans les cas que nous avons rencontrés, la relation se simplifie et devient :

$$V(a_i) = \frac{1}{n} V(e)$$

La variance des coefficients est donc directement liée à la variance de l'écart. Rappelons que l'écart peut être divisé en deux parties : le *manque d'ajustement* (*lack of fit* en anglais) et l'*erreur expérimentale* (*pure error* en anglais) :

$$e = + \Delta + \varepsilon$$

Et que l'on a :

$$V(e) = V(\Delta) + V(\varepsilon)$$

Pour obtenir la grandeur étalon que nous recherchons, il faut posséder une valeur de la variance de l'écart. On obtient cette valeur grâce à l'analyse de la variance.

5.4 Analyse de la variance

5.4.1 Principes de l'analyse de la variance

L'analyse de la variance consiste à rechercher les sources de variation des réponses. On suppose que les réponses ont été calculées avec le modèle postulé, $y_i = f(x_1, x_2, x_3, \dots, x_n) + e_i$, en utilisant la méthode des moindres carrés c'est-à-dire en minimisant la somme des carrés des écarts. Dans ce cas, les réponses calculées s'écrivent \hat{y}_i et les écarts e prennent des valeurs particulières r_i qui s'appellent les *résidus*. Les résidus sont donc des valeurs particulières des écarts. On a :

$$\hat{y}_i = f(x_1, x_2, x_3, \dots, x_n)$$

Avec ces nouvelles notations, la relation donnant la réponse peut s'écrire :

$$y_i = \hat{y}_i + r_i$$

L'analyse classique de la variance fait intervenir non pas les réponses mais la différence entre les réponses et leur moyenne, $(y_i - \bar{y})$ ou $(\hat{y}_i - \bar{y})$. Cette différence est désignée soit par *écarts à la moyenne*, soit par *réponse corrigée de la moyenne*. Dans le cas des réponses calculées, on parle aussi de *modèle corrigé de la moyenne*.

Dans le cadre de la méthode des moindres carrés, la moyenne des réponses mesurées est égale à la moyenne des réponses calculées avec le modèle postulé. On a donc, \bar{y} étant la moyenne des réponses :

$$y_i - \bar{y} = \hat{y}_i - \bar{y} + r_i$$

Lorsqu'on élève les deux membres de cette relation au carré, on obtient :

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum r_i^2 \quad \{5.2\}$$

C'est la relation de base de l'analyse de la variance. Le membre de gauche est la somme des carrés des écarts à la moyenne des réponses mesurées. Cette somme se décompose en deux éléments : la somme des carrés des écarts à la moyenne des réponses calculées avec le modèle et la somme des carrés des résidus.

La somme des carrés des résidus est la plus faible valeur de la somme des carrés des écarts. On a donc :

$$\sum r_i^2 = \text{Minimum de } \sum e_i^2 = \text{Minimum de } \sum (\Delta + \sigma)^2$$

Si l'on divise la somme des carrés des résidus par le nombre de degrés de liberté des résidus, on obtient la variance des *résidus*. La variance des résidus $V(r_i)$ est donc la plus petite variance des écarts $V(e)$. On peut écrire :

$$V(r_i) = \text{Minimum de } V(e) = \frac{1}{n-p} \sum_{i=1}^{i=n} r_i^2$$

C'est cette valeur minimale de la variance des écarts qui est généralement adoptée comme étalon de comparaison pour évaluer l'importance d'un coefficient. La variance des coefficients est alors calculée :

– soit par la formule générale utilisée par les ordinateurs :

$$V(a_i) = K V(e) = V(r_i)$$

– soit par la formule simplifiée lorsqu'il s'agit de plans factoriels et d'un modèle postulé polynomial :

$$V(a_i) = \frac{1}{n} V(e) = \frac{1}{n} V(r_i) \quad \{5.3\}$$

En résumé, la variance des résidus de l'analyse de la variance sert à calculer la variance des coefficients. C'est la variance des coefficients qui sert d'étalon pour tester si un coefficient est significatif ou non.

Remarque 1

Les n réponses mesurées sont toutes indépendantes, c'est-à-dire qu'il n'existe pas de relation mathématique entre elles. Elles possèdent donc n degrés de liberté. On a utilisé un degré de liberté pour calculer la moyenne. La variance des écarts à la moyenne des réponses mesurées, $y_i - \bar{y}$, possède $n - 1$ degrés de liberté.

Le calcul des p coefficients a pris $p - 1$ degrés de liberté seulement car la moyenne avait déjà été calculée. La variance des écarts à la moyenne des réponses calculées, $\hat{y}_i - \bar{y}$, possède donc $p - 1$ degrés de liberté.

Il reste donc $n - 1 - (p - 1) = n - p$ degrés de liberté pour le calcul de la variance des résidus.

Remarque 2

Les règles et les tests statistiques ne s'appliquent qu'aux variables purement aléatoires. Le manque d'ajustement n'étant pas une grandeur aléatoire mais une erreur systématique, on ne devrait pas appliquer les règles statistiques aux écarts. On le fait pourtant car on suppose que le manque d'ajustement est de l'ordre de grandeur de l'erreur expérimentale. Il est donc bon de s'assurer de la validité de cette hypothèse : le manque d'ajustement est de l'ordre de grandeur de l'erreur expérimentale. La mesure de l'erreur expérimentale $V(\sigma)$ est assurée par des répétitions. La mesure de la variance des résidus $V(r)$ est tirée de l'analyse de variance. On peut donc calculer la variance de l'ajustement $V(\Delta)$:

$$V(r) = V(\Delta) + V(\sigma)$$

En comparant la variance de l'erreur expérimentale à la variance de l'ajustement, il est possible de voir si l'hypothèse est valide.

Remarque 3

Il ne faut pas confondre les écarts, les résidus et les écarts à la moyenne.

Les écarts sont les différences, e , entre les réponses mesurées et le modèle mathématique postulé avant d'effectuer les calculs des coefficients par la méthode des moindres carrés :

$$y_i = f(x_1, x_2, x_3, \dots, x_n) + e_i$$

Les résidus sont les différences, r_i , entre les réponses mesurées et les réponses calculées avec des coefficients obtenus par la méthode des moindres carrés :

$$r_i = y_i - \hat{y}_i$$

Les écarts à la moyenne sont les différences $y_i - \bar{y}$ ou $\hat{y}_i - \bar{y}$ entre les réponses et la moyenne des réponses.

Remarque 4

Le carré moyen des résidus est obtenu en divisant la somme des carrés des résidus par le nombre de degrés de liberté correspondant. C'est donc une grandeur analogue à une variance. Par conséquent, la racine carrée de cette variance est analogue à un écart-type. C'est pourquoi, on trouve souvent les expressions *variance résiduelle* et *écart-type des résidus* pour qualifier ces grandeurs. Mais comme les écarts ne sont pas entièrement aléatoires, certains auteurs préfèrent utiliser le terme *carré moyen des résidus* à la place de variance résiduelle et *racine carrée du carré moyen des résidus* au lieu d'écart-type des résidus. L'expression *racine carrée du carré moyen des résidus* étant un peu longue, on a l'habitude d'utiliser à la place l'abréviation anglo-saxonne RMSE (*Root Mean Square Error*).

Quand le modèle est bien ajusté, on utilise cette RMSE pour calculer l'erreur sur les coefficients du modèle postulé.

Quelles que soient les dénominations utilisées pour ces grandeurs, elles impliquent de longs calculs. Heureusement les ordinateurs réalisent tous ces calculs et fournissent directement les résultats de l'analyse de la variance sous la forme d'un tableau. Le seul élément choisi par l'expérimentateur est le modèle *a priori* qui sert à obtenir les réponses calculées. Les résultats de l'analyse de la variance seront dépendants du choix du modèle. Nous invitons les lecteurs à refaire les calculs de l'analyse de la variance en utilisant différents modèles postulés. Ils verront l'influence du choix du modèle sur la valeur de la RMSE.

5.4.2 Présentation de l'analyse de la variance

Les logiciels, et même les tableurs, ont des programmes qui construisent les tableaux d'analyse de la variance. Les plus simples de ces tableaux comportent 5 colonnes (source de variation, somme des carrés, degrés de liberté, carré moyen et statistique de Fisher) et 4 lignes (titres, modèle corrigé de la moyenne, résidus et réponses mesurées corrigées de la moyenne) semblables au tableau 5.5. La première colonne indique les sources de variation. On y trouve le modèle, les résidus et les réponses mesurées. La deuxième colonne donne les sommes des carrés des écarts à la moyenne. On remarque que la somme des carrés des réponses mesurées (corrigées de la moyenne) est égale à la somme des deux autres sommes de carrés. La troisième colonne indique les degrés de liberté (ddl) de chaque somme de carrés. On remarque également que la somme des ddl du modèle et des résidus est égale aux ddl des réponses mesurées. Les *carrés moyens* de la quatrième colonne sont les sommes des carrés divisées par les ddl. On notera que la racine carrée du carré moyen des résidus sert à calculer l'étalon permettant de tester les coefficients. C'est donc une statistique très importante. Enfin, la cinquième colonne indique le *F de Fisher* qui est le rapport du carré moyen du modèle à celui des résidus. Ce rapport permet de calculer la probabilité que ces deux carrés ne soient pas égaux. En d'autres termes, si le F de Fisher est élevé (faible probabilité que le modèle soit celui de la moyenne), les variations des réponses mesurées sont bien dues aux variations des facteurs. Si le F de Fisher est proche de 1 (forte probabilité que le modèle soit celui de la moyenne), les variations des réponses mesurées sont comparables à celles des résidus.

Tableau 5.5 – Analyse de la variance.

Source des variations	Somme des carrés	ddl	Carré moyen	Fisher
Modèle corrigé de la moyenne	965,30	10	96,530	46,3196
Résidus	10,42	5	2,084	Prob F 0,0003
Réponses mesurées corrigées de la moyenne	975,72	15		

■ Coefficient de détermination R^2

L'analyse de la variance permet de calculer une statistique très utile : le R^2 ou R carré. Cette statistique est le rapport de la somme des carrés des réponses calculées (corrigées de la moyenne) à la somme des carrés des réponses mesurées (corrigées de la moyenne) :

$$R^2 = \frac{\text{Somme des carrés des réponses calculées corrigées de la moyenne}}{\text{Somme des carrés des réponses mesurées corrigées de la moyenne}}$$

Dans cet exemple :

$$R^2 = \frac{965,30}{975,72} = 0,9893$$

Si le modèle permet de retrouver exactement la valeur des réponses mesurées, la somme des carrés des réponses calculées est égale à la somme des carrés des réponses mesurées. Le R^2 est égal à 1.

Si le modèle fournit des réponses calculées égales à la moyenne, la somme des carrés des réponses calculées est égale à 0. Le R^2 est égal à 0. On dit que le modèle n'a pas de puissance d'explication.

Le R^2 est donc une mesure de la qualité du modèle. S'il est égal à 1, le modèle permet de retrouver la valeur des réponses mesurées. S'il est égal à 0, le modèle n'en dit pas plus que la moyenne des réponses. Dans la pratique, il est difficile d'indiquer la valeur d'un bon R^2 car les valeurs varient beaucoup d'une discipline à l'autre. Par exemple, un R^2 de 0,8 peut être considéré comme mauvais par un ingénieur et un R^2 de 0,3 peut être vu comme bon par un psychologue.

Le R^2 joue véritablement son rôle d'indicateur de la qualité du modèle à condition que les résidus ne soient pas nuls. Si les résidus sont nuls, le R^2 est égal à 1 quelle que soit la qualité du modèle. On observe un tel cas lorsqu'il n'y a pas de degrés de liberté. Le R^2 est donc un bon indicateur de la qualité du modèle s'il y a plus de points expérimentaux différents que de coefficients dans le modèle postulé.

5.5 Application à l'exemple 04

5.5.1 Analyse de la variance

Nous allons construire la table de l'analyse de la variance de l'exemple sur les outils jetables et en déduire l'importance de chaque coefficient. Rassemblons d'abord les

données de base : les réponses mesurées, les réponses calculées avec un modèle et les résidus correspondants (Tableau 5.6). Le modèle choisi est le modèle des plans factoriels avec les interactions d'ordre 2. Il contient 11 coefficients :

$$\hat{y} = 11,55 - 0,1x_1 - 6,2875x_2 - 3,425x_3 - 2,075x_4 \\ + 0,6375x_1x_2 + 0,325x_1x_3 + 0,35x_1x_4 + 1,5875x_2x_3 - 1,1125x_2x_4 + 0,6x_3x_4$$

Tableau 5.6 – Données de base de l'analyse de la variance.

Essai n°	Réponses mesurées	Réponses calculées	Résidus
1	26,1	25,825	0,275
2	22,2	23	-0,8
3	10,1	11,025	-0,925
4	12,2	10,75	1,45
5	14,2	13,95	0,25
6	12,7	12,425	0,275
7	5,9	5,5	0,4
8	5,6	6,525	-0,925
9	23	22	1
10	20,1	20,575	-0,475
11	2,4	2,75	-0,35
12	3,7	3,875	-0,175
13	11	12,525	-1,525
14	13,4	12,4	1
15	0,5	-0,375	0,875
16	1,7	2,05	-0,35

■ Somme des carrés des réponses mesurées corrigées de la moyenne

La moyenne des réponses mesurées est 11,55.

1. On commence par calculer les écarts à la moyenne des 16 réponses mesurées au cours du plan factoriel :

$$\begin{aligned} 26,1 - 11,55 &= +14,55 \\ 22,2 - 11,55 &= +10,65 \\ 10,1 - 11,55 &= -1,45 \\ &\dots \\ 0,5 - 11,55 &= -11,05 \\ 1,7 - 11,55 &= -9,85 \end{aligned}$$

2. Puis, ces écarts à la moyenne sont élevés au carré et additionnés. On obtient ainsi la somme des carrés des écarts à la moyenne des réponses mesurées corrigées de la moyenne :

$$(14,55)^2 + (10,65)^2 + \dots + (-11,05)^2 + (-9,85)^2 \\ = + 211,702\ 5 + 113,422\ 5 + \dots + 122,102\ 5 + 97,022\ 5 = + 975,72$$

■ **Carré moyen des réponses calculées avec le modèle et corrigées de la moyenne (variance des réponses calculées ou variance du modèle)**

1. On calcule les 16 réponses avec le modèle défini plus haut. On vérifie que leur moyenne est égale à 11,55.

2. On calcule les écarts à la moyenne des 16 réponses calculées :

$$25,825 - 11,55 = +14,275 \\ 23,0 - 11,55 = +11,45 \\ 11,025 - 11,55 = -0,525 \\ \dots \\ -0,375 - 11,55 = -11,925 \\ 2,05 - 11,55 = -9,5$$

3. Ces écarts à la moyenne sont élevés au carré et additionnés. On obtient ainsi la somme des carrés des écarts à la moyenne des réponses calculées corrigées de la moyenne :

$$(14,275)^2 + (11,45)^2 + \dots + (-11,925)^2 + (-9,5)^2 \\ = + 203,775\ 625 + 131,102\ 5 + \dots + 142,205\ 625 + 90,25 = + 965,30$$

4. Cette somme est divisée par le nombre de coefficients du modèle moins 1. On obtient le carré moyen des réponses calculées avec le modèle (ou variance des réponses calculées ou variance du modèle) :

$$\frac{965,30}{11 - 1} = 96,53$$

■ **Carré moyen des résidus (variance des résidus)**

Rappelons qu'un résidu est égal à :

$$r_i = y_i - \hat{y}_i$$

1. On calcule les 16 résidus :

$$26,1 - 25,825 = +0,275 \\ 22,2 - 23 = -0,8 \\ 10,1 - 11,025 = -0,925 \\ \dots \\ 0,5 - (-0,375) = -0,875 \\ 1,7 - 2,05 = -0,35$$

2. Ces résidus sont élevés au carré et additionnés. On obtient ainsi la somme des carrés des résidus :

$$+0,075\ 625 + 0,64 + \dots + 0,765\ 625 + 0,122\ 5 = +10,42$$

3. Cette somme est divisée par $n - p = 16 - 11 = 5$ degrés de liberté des résidus :

$$\frac{10,42}{16 - 11} = 2,084$$

On obtient ainsi le carré moyen des résidus. La racine carrée de ce carré moyen est la RMSE qui sert d'étalon pour évaluer l'importance des coefficients. La RMSE vaut ici 1,4436 lorsque le modèle postulé ne contient que les effets principaux et leurs interactions d'ordre 2.

■ Écart-type des coefficients

Dans le cas de l'exemple des outils de coupe jetables, il s'agit d'un plan factoriel où les niveaux -1 et $+1$ ont été respectés. Le modèle est bien un modèle polynomial, on peut donc appliquer la formule simplifiée et obtenir ainsi l'étalon de comparaison. Chaque coefficient a été calculé avec $n = 16$ réponses :

$$V(a_i) = \frac{1}{n}V(r_i) = \frac{1}{16}2,084 = 0,130\ 25$$

$$\sigma_{a_i} = \sqrt{0,130\ 25} = 0,360\ 9$$

C'est la valeur de comparaison que nous cherchions.

■ p -value

La p -value est la probabilité qu'un coefficient soit négligeable. On calcule cette p -value à partir du rapport du coefficient à son écart-type (t de Student). Par exemple, le t de Student du coefficient du débit est $-0,1/0,360\ 9 = 0,28$, et la p -value associée est égale à 0,792 8.

Il est maintenant possible de construire un tableau où chaque coefficient figure accompagné de sa valeur, de son écart-type, du t de Student et de la p -value (Tableau 5.7). Ce tableau est identique au tableau 5.3. La seule différence est que maintenant vous êtes à même de comprendre la signification de tous les chiffres.

La constante à moins de 1 chance sur 10 000 d'être égale à 0. Elle n'est donc certainement pas égale à 0, elle est significative et il faut la conserver dans le modèle final.

Le débit (1) à une p -value de 0,79, il a donc une forte chance d'être égal à 0. Il peut être considéré comme nul ou comme très petit. Il n'est donc pas significatif et on peut l'éliminer du modèle final. Cela signifie que les réponses calculées ne seront pas très différentes, que l'on utilise un modèle final avec ou sans ce coefficient.

Si l'on fixe la probabilité d'acceptation des coefficients à une p -value de 0,10, on conservera la vitesse de coupe (2), la profondeur de coupe (3), la vitesse d'avance (4) ainsi que les interactions 23 et 24. Le modèle final s'écrit alors :

$$y = 11,55 - 6,29x_2 - 3,42x_3 - 2,1x_4 + 1,6x_2x_3 - 1,1x_2x_4 \quad \{5.4\}$$

On vérifie que ce modèle possède un bon R^2 (0,97) et que les p -values des coefficients sont toujours inférieures à 0,1. Si l'on avait fixé la probabilité d'acceptation des coefficients à un autre niveau, on aurait pu obtenir un autre modèle final. Le choix du modèle définitif est laissé à l'appréciation de l'expérimentateur qui, lui, connaît les enjeux et les risques de l'étude.

Tableau 5.7 – Importance des coefficients du modèle (unités codées)

Coefficient	Valeur	Écart-type	t de Student	p-value
Constante	11,55	0,360 9	32	< 0,000 1
Débit (1)	-0,1	0,360 9	-0,28	0,792 8
Vitesse de coupe (2)	-6,287 5	0,360 9	-17,42	< 0,000 1
Profondeur (3)	-3,425	0,360 9	-9,49	0,000 2
Vitesse d'avance (4)	-2,075	0,360 9	-5,75	0,002 2
Interaction (12)	0,637 5	0,360 9	1,77	0,137 6
Interaction (13)	0,325	0,360 9	0,90	0,409 1
Interaction (14)	0,35	0,360 9	0,97	0,376 7
Interaction (23)	1,587 5	0,360 9	4,40	0,007 0
Interaction (24)	-1,112 5	0,360 9	-3,08	0,027 4
Interaction (34)	0,6	0,360 9	1,66	0,157 3

5.5.2 Analyse des résidus

Si l'interprétation a été bien faite, le modèle doit bien représenter les réponses mesurées (R^2 proche de 1) et la racine carrée du carré moyen des résidus doit être du même ordre de grandeur que l'erreur expérimentale. On peut encore s'assurer qu'il ne reste pas un peu d'information structurée dans les résultats en examinant le graphique des résidus. Il est donc important de regarder comment les résidus se répartissent en fonction de différents critères. S'ils semblent répartis au hasard, il ne reste probablement plus rien à extraire des données. C'est le cas de l'exemple de ce chapitre (Figure 5.9).

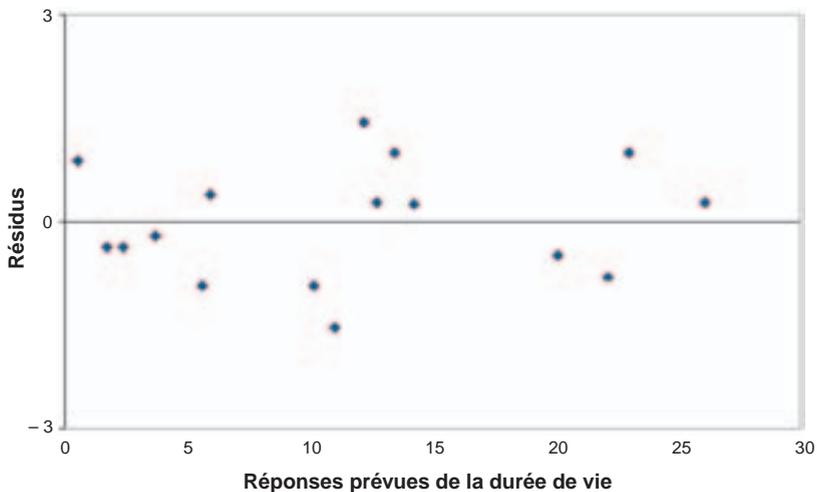


Figure 5.9 – Analyse des résidus de l'exemple sur les outils de coupe.

En revanche, si l'analyse des résidus fait apparaître des dispositions remarquables, il reste un peu d'information à extraire. Par exemple, on peut rencontrer des diagrammes où les résidus sont de plus en plus élevés lorsque la réponse augmente (Figure 5.10 haut). Dans ce cas, l'hypothèse d'homoscédasticité (même erreur expérimentale dans tout le domaine d'étude) n'est pas vérifiée. On peut aussi rencontrer des diagrammes où les résidus sont négatifs pour les faibles et les fortes valeurs de la réponse et positifs pour les valeurs intermédiaires (Figure 5.10 bas). Dans ce cas, l'hypothèse d'un modèle simple du premier degré ou d'un modèle du premier degré avec interactions (modèle PDAI) est probablement à remettre en cause et il faut passer à un modèle du second degré.

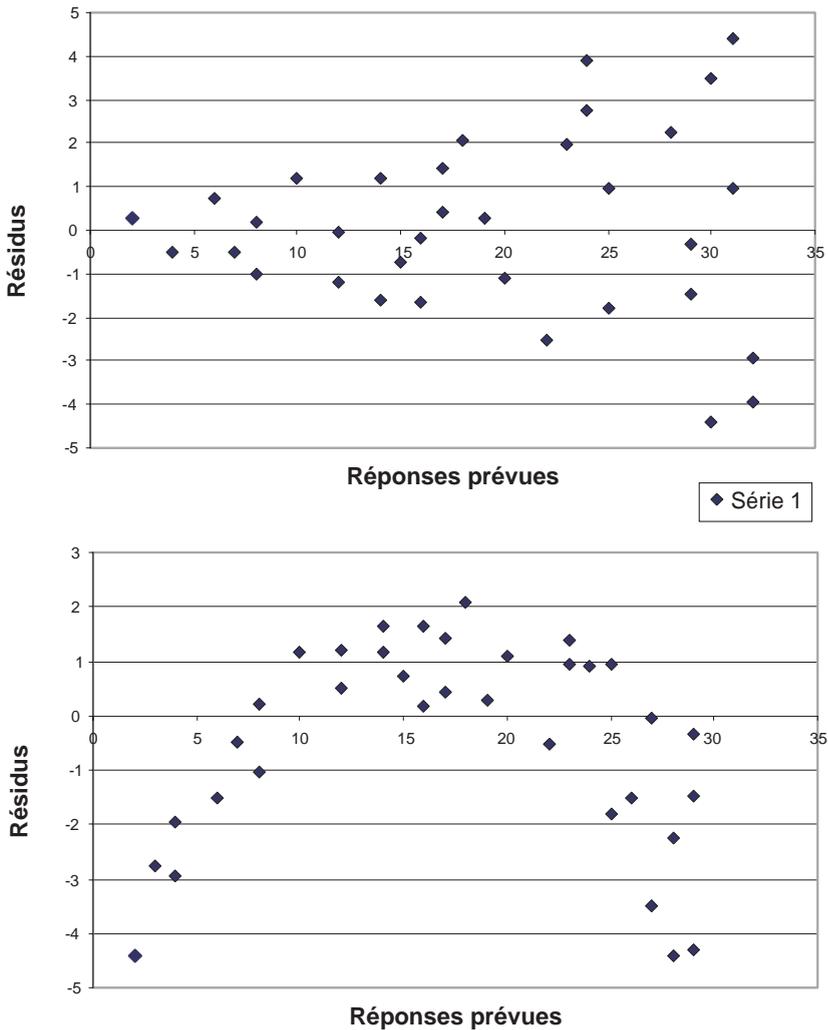


Figure 5.10 – Exemples d'analyse des résidus montrant une structure particulière : en haut, la dispersion des résidus augmente avec la réponse (il n'y a pas homoscédasticité), en bas, les résidus laissent supposer un modèle du second degré.

5.6 Conclusion de l'étude

La durée de vie des outils de coupe est surtout influencée par la vitesse et la profondeur de coupe. Pour ménager les outils, il faut travailler à faible vitesse de coupe et ne pas prendre des passes trop profondes. Mais si l'on veut assurer la rentabilité du procédé, il faut que le rendement soit suffisamment élevé. Il faudra choisir le compromis qui assure une production rentable à une usure contrôlée des outils de coupe. Ce compromis est atteint si la durée de vie est au moins de 20 heures. Le modèle établi (unités codées) permet de faire des prévisions :

$$y = 11,55 - 6,29x_2 - 3,42x_3 - 2,1x_4 + 1,6x_2x_3 - 1,1x_2x_4$$

Le diagramme de la figure 5.11, tracé pour un débit de 725 et une vitesse d'avance de 0,75, indique une zone potentiellement favorable qui se situe dans le quart inférieur gauche, c'est-à-dire aux faibles vitesses de coupe et aux faibles profondeurs. D'après les résultats du plan d'expériences, les essais n° 1, 2, 9 et 10 répondent à ces impératifs et on vérifie que les durées de vie sont supérieures à 20 heures (Tableau 5.8).

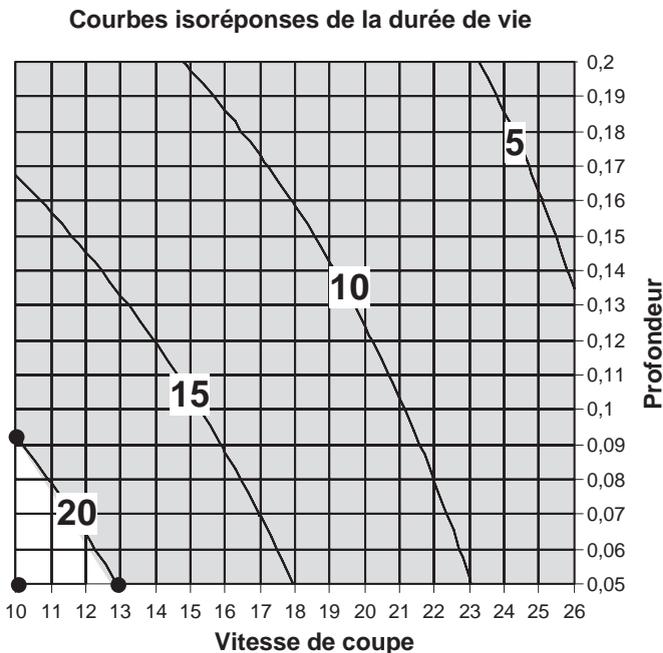


Figure 5.11 – Courbes isoréponses de la durée de vie des outils de coupe en fonction de la vitesse de coupe et de la profondeur. Le débit et la vitesse d'avance sont fixés respectivement à 725 et à 0,75.

Mais avant de faire des propositions, il est bon de réaliser des expériences de contrôle pour confirmer les prévisions. On réalise donc trois essais (Tableau 5.9), l'un aux bornes du domaine (essai n° 21) et deux aux limites des 20 heures (essais n° 22 et 23).

Tableau 5.8 – Essais dont les durées de vie sont supérieures à 20 heures.

Essai n°	Débit (1)	Vitesse de coupe (2)	Profondeur (3)	Vitesse d'avance (4)	Réponses
1	650	10	0,05	0,5	26,1
2	800	10	0,05	0,5	22,2
9	650	10	0,05	1	23
10	800	10	0,05	1	20,1

Tableau 5.9 – Essais de vérification.

Essai n°	Débit (1)	Vitesse de coupe (2)	Profondeur (3)	Vitesse d'avance (4)	Réponses prévues	Réponses mesurées
21	725	10	0,05	0,75	22,86	23,1
22	725	13	0,05	0,75	19,90	20,2
23	725	10	0,09	0,75	20,18	19,8

On compare les prévisions faites avec le modèle simplifié {5.4} aux mesures effectuées. On constate un bon accord. Le modèle est donc validé et les recommandations pourraient être les suivantes.

Afin de préserver les outils de coupe et leur assurer une durée de vie proche de 20 heures ou plus, il ne faut pas dépasser une vitesse de coupe de 11, la profondeur de coupe doit être inférieure à 0,07 et l'avance doit être plus petite que 0,75 (Figure 5.12). Ces recommandations sont faciles à trouver avec la fonction désirabilité (chapitre 4).

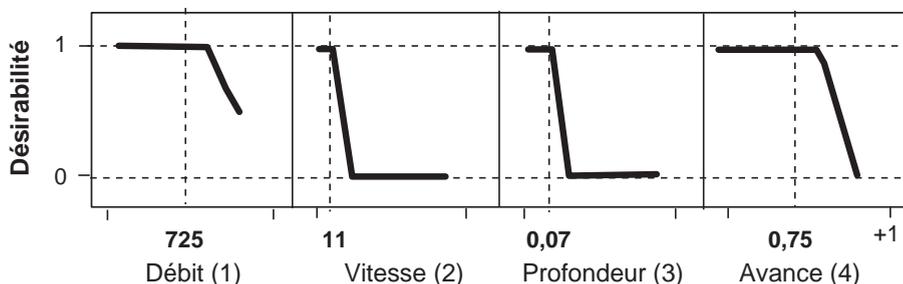


Figure 5.12 – La désirabilité permet de trouver rapidement les bonnes conditions opératoires.

6 • PLANS FACTORIELS FRACTIONNAIRES

Comme les plans factoriels complets, les plans factoriels fractionnaires possèdent des facteurs ayant chacun deux niveaux, un niveau bas et un niveau haut. Mais, on ne réalise pas toutes les combinaisons de niveaux. En effet, le nombre d'essais des plans factoriels complets augmente rapidement avec le nombre de facteurs. Il n'est pas rare de rencontrer des études faisant intervenir 7 ou 8 facteurs. Il est bien sûr déraisonnable de vouloir exécuter $2^7 = 128$ ou $2^8 = 256$ essais. On sélectionne une fraction des essais d'un plan complet pour construire un plan fractionnaire. Cette sélection est basée sur des considérations mathématiques qui seront mises en application grâce au calcul de Box. Il est aussi possible d'utiliser les logiciels de plans d'expériences qui mettent en œuvre ces principes mathématiques. L'interprétation des plans fractionnaires est moins facile que celle des plans complets puisqu'on possède moins d'information. On se contente souvent de savoir quels sont les facteurs influents, et parfois on va un peu plus loin en examinant les interactions d'ordre 2. C'est la raison pour laquelle ces plans sont souvent appelés *plans de criblage*. Ces plans permettent d'étudier beaucoup de facteurs et de déterminer ceux qui sont influents. L'analyse ne va souvent pas plus loin et atteint rarement la modélisation mathématique des réponses.

Les modèles mathématiques utilisés sont donc, eux aussi, beaucoup plus simples. On emploie des modèles ne comportant que les facteurs principaux au premier degré ou des modèles avec des interactions d'ordre 2 suivant les objectifs de l'étude.

Au lieu de réaliser un plan complet 2^4 de 16 essais, on peut n'exécuter que 8 essais, c'est-à-dire la moitié des essais. Le plan fractionnaire correspondant est un plan 2^4 divisé par 2, soit un plan $2^4/2$ ou 2^{4-1} . La signification des termes de la dénomination 2^{4-1} est la suivante : le 2 indique que les facteurs prennent chacun deux niveaux, le 4 signifie que l'on étudie 4 facteurs. Enfin, $4 - 1 = 3$ signifie que l'on utilise un plan analogue au plan factoriel complet 2^3 qui ne comprend que 8 essais. On remarque également que cette manière de désigner un plan indique le nombre d'essais à effectuer puisque l'on a $2^{4-1} = 8$.

L'interprétation des plans fractionnaires nécessite de connaître la théorie des alias et d'appliquer des hypothèses d'interprétation. La première des principales hypothèses d'interprétation est la suivante :

Hypothèse 1 – Les interactions d'ordre trois et toutes celles d'ordre plus élevé peuvent être négligées.

6.1 Exemple 05 : le tellure

6.1.1 Préparation du plan d'expériences

■ Présentation de l'étude sur le tellure

Des chimistes cherchent à déterminer la teneur en tellure (Te) dans l'eau de mer. Mais la nature et la concentration d'autres métaux faussent les mesures. La teneur en tellure peut être soit trop élevée, soit trop faible en fonction des autres métaux et de leur concentration. Les métaux perturbateurs pourraient être le sodium (Na), le potassium (K), le calcium (Ca) et le magnésium (Mg).

■ Définition de l'objectif de l'étude

Les chimistes désirent savoir quels sont les métaux qui perturbent leur mesure et avoir une idée de l'importance des différences entre la concentration mesurée et la concentration réelle en tellure. Ils veulent pouvoir répondre à la question : « La teneur en tellure mesurée est-elle beaucoup, moyennement ou peu perturbée par la présence de Na, de K, de Ca et de Mg ? » Pour réaliser cette étude, les chimistes préparent des solutions contenant toujours la même quantité de tellure auxquelles ils ajoutent des teneurs variables en métaux perturbateurs.

Il ne s'agit pas de faire des corrections de mesure en fonction des teneurs en Na, K, Ca et Mg. Dans ce cas, il faudrait faire un plan complet de 16 essais éventuellement suivi d'un plan du second degré si le modèle PDAI était insuffisant. Il s'agit simplement de connaître l'importance de l'erreur provoquée par la présence des quatre métaux testés.

■ Choix de la réponse permettant d'atteindre l'objectif

La réponse choisie est la différence entre la concentration vraie en tellure et la concentration déterminée par spectrométrie atomique.

■ Recherche des facteurs qui pourraient être influents sur la réponse

Pour les quatre métaux perturbateurs on retient :

- la concentration en sodium,
- la concentration en potassium,
- la concentration en calcium,
- la concentration en magnésium.

Le tableau 6.1 indique les concentrations et le domaine d'étude. On notera que le point central n'est pas du tout au milieu de l'intervalle [2,5 ; 250]. La raison est que les chimistes savent qu'il faut faire intervenir les concentrations par leur logarithme. Avec une échelle logarithmique le niveau 0 est bien au point central.

Les facteurs étudiés et qui seront utilisés dans les modèles mathématiques sont :

- le logarithme de la concentration en sodium Na,
- le logarithme de la concentration en potassium K,
- le logarithme de la concentration en calcium Ca,
- le logarithme de la concentration en magnésium Mg.

Tableau 6.1 – Facteurs retenus et domaine d'étude.

Facteurs	Niveau -1 ($\mu\text{g/ml}$)	Niveau 0 ($\mu\text{g/ml}$)	Niveau +1 ($\mu\text{g/ml}$)
Concentration Na (1)	2,5	25	250
Concentration K (2)	2,5	25	250
Concentration Ca (3)	2,5	25	250
Concentration Mg (4)	2,5	25	250

■ Choix du plan d'expériences

L'expérimentateur pourrait faire un plan complet 2^4 , soit 16 essais. Il pourrait ainsi évaluer les effets principaux, les interactions entre 2, 3 et 4 facteurs. Mais d'après ses connaissances, il est peu probable qu'il y ait des interactions entre les métaux perturbateurs. Il décide donc de faire un plan de 8 essais, c'est-à-dire un plan 2^{4-1} . Il pourra ainsi évaluer les effets principaux de chaque facteur et savoir si des interactions d'ordre 2 risquent d'être influentes.

Le tableau 6.2 indique les détails du plan d'expériences et les réponses mesurées au cours de l'expérimentation.

6.1.2 Expérimentation

Tableau 6.2 – Essais et réponses du plan d'expériences sur le tellure.

Essai n°	Na (1)	K (2)	Ca (3)	Mg (4)	Réponse
1	-1	-1	-1	-1	128
2	+1	-1	-1	+1	153
3	-1	+1	-1	+1	104
4	+1	+1	-1	-1	130
5	-1	-1	+1	+1	86
6	+1	-1	+1	-1	109
7	-1	+1	+1	-1	66
8	+1	+1	+1	+1	96
-1	Log 2,5	Log 2,5	Log 2,5	Log 2,5	
0	Log 25	Log 25	Log 25	Log 25	
+1	Log 250	Log 250	Log 250	Log 250	

6.1.3 Interpréter les résultats des calculs

Pour interpréter les résultats expérimentaux, on peut traiter les données brutes de plusieurs manières. On peut éliminer des coefficients du modèle mathématique pour avoir autant (ou moins) d'inconnues que d'équations. On peut aussi regrouper les coefficients pour avoir, là aussi, autant (ou moins) d'inconnues que d'équations. Dans ce dernier cas, il faut utiliser la théorie des alias pour comprendre ce que signifient les valeurs numériques calculées. On peut aussi employer une méthode mixte basée à la fois sur des modèles simplifiés et sur la théorie des alias.

Cette interprétation des résultats est facilitée par les hypothèses d'interprétation. Nous verrons plusieurs plans fractionnaires et nous verrons comment des raisonnements simples permettent, soit de conclure que l'étude est finie, soit de décider qu'il faut faire des essais supplémentaires pour lever les difficultés rencontrées.

■ Modèle mathématique

En accord avec la première hypothèse d'interprétation (« Les interactions d'ordre trois et toutes celles d'ordre plus élevé peuvent être négligées »), on choisit un modèle faisant intervenir tous les termes du premier degré et toutes les interactions entre deux facteurs :

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{14}x_1x_4 + a_{23}x_2x_3 + a_{24}x_2x_4 + a_{34}x_3x_4 \quad \{6.1\}$$

Il y a donc 11 coefficients. Le logiciel calculera le coefficient constant, les effets principaux et trois groupes d'interactions. Si les groupes d'interactions calculées sont faibles, on pourra les éliminer. Si les groupes d'interactions calculées sont forts, on ne pourra pas les éliminer et l'interprétation deviendra plus délicate et il faudra faire appel à la théorie des alias et à de nouvelles hypothèses d'interprétation.

Cette modélisation est un simple outil permettant de connaître rapidement les facteurs qui risquent d'avoir le plus d'influence sur le phénomène étudié. Il ne s'agit pas d'une modélisation permettant de faire des prévisions.

Tableau 6.3 – Coefficients du modèle (unités codées).

Effet	Valeur
Constante	109
Na (1)	13
K (2)	-10
Ca (3)	-19,75
Mg (4)	0,75
Na K (12) + Ca Mg (34)	1
Na Ca (13) + K Mg (24)	0,25
K Ca (23) + Na Mg (14)	1,75

On constate (Tableau 6.3) qu'un facteur a un effet particulièrement faible : le facteur 4 (magnésium). La teneur en tellure est faussée par les trois autres métaux : le sodium (Na), le potassium (K) et le calcium (Ca). Les sommes d'interactions sont moins élevées que les effets des facteurs principaux et peuvent être considérées comme faibles.

6.2 Théorie des aliases

La *théorie des aliases* est très utile pour comprendre et interpréter les résultats d'un plan fractionnaire. Nous avons divisé l'étude de cette théorie en cinq parties :

1. Définition des contrastes.
2. Hypothèses d'interprétation.
3. Calcul de Box.
4. Relation d'équivalence.
5. Générateurs d'aliases.

À la suite de la théorie des aliases nous verrons deux aspects importants des plans fractionnaires :

- la construction des plans fractionnaires,
- le nombre maximal de facteurs étudiés sur un plan de base donné.

6.2.1 Définition des contrastes

Lors de l'interprétation des résultats d'un plan fractionnaire, on applique les mêmes règles de calcul que celles des plans complets. Le modèle postulé pour analyser les résultats de l'exemple sur le tellure est alors :

$$\begin{aligned}
 y = & a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 \\
 & + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{14}x_1x_4 + a_{23}x_2x_3 + a_{24}x_2x_4 + a_{34}x_3x_4 \\
 & + a_{123}x_1x_2x_3 + a_{124}x_1x_2x_4 + a_{134}x_1x_3x_4 + a_{234}x_2x_3x_4 \\
 & + a_{1234}x_1x_2x_3x_4
 \end{aligned} \quad \{6.2\}$$

Ce modèle comporte 16 coefficients inconnus. Pour les calculer, il faut posséder 16 équations. C'est bien le cas avec les plans complets. Mais, avec le plan fractionnaire employé, il n'y a que 8 équations. La matrice du modèle comporte donc 8 lignes et 16 colonnes. On s'arrange pour que les colonnes possèdent, deux à deux, des colonnes ayant des niveaux identiques. Les inconnues sont ainsi regroupées deux par deux. On obtient un système de 8 groupes d'inconnues pour 8 équations. Dans l'exemple du tellure, les niveaux du facteur 4 sont les mêmes que les niveaux de l'interaction 123, on a donc la colonne du facteur 4 identique à la colonne de l'interaction 123. Pour ce plan, on peut écrire $x_4 = x_1x_2x_3$. On verra dans les paragraphes suivants comment s'associent les colonnes et par conséquent les coefficients en fonction des points d'expériences retenus pour un plan fractionnaire donné. On possède de cette manière 8 groupes de 2 inconnues et on peut résoudre le système. Ces regroupements de coefficients s'appellent des *aliases* ou des *contrastes*. On les note par la lettre ℓ_i . Dans l'exemple du tellure, le contraste ℓ_1 regroupe le coefficient a_1 du facteur 1 et le coefficient a_{234} de l'interaction 234. On dit que le coefficient a_1 est *aliasé* avec l'interaction a_{234} .

De même :

- Le coefficient a_2 est aliasé avec l'interaction a_{134} .

- Le coefficient a_3 est alié avec l'interaction a_{124} .
- Le coefficient a_4 est alié avec l'interaction a_{123} .
- Le coefficient a_{12} est alié avec l'interaction a_{34} .
- Le coefficient a_{13} est alié avec l'interaction a_{24} .
- Le coefficient a_{23} est alié avec l'interaction a_{14} .
- Le coefficient a_0 est alié avec l'interaction a_{1234} .

La relation {6.2} s'écrit alors :

$$y = (a_0 + a_{1234}) + (a_1 + a_{234})x_1 + (a_2 + a_{134})x_2 + (a_3 + a_{124})x_3 + (a_4 + a_{123})x_4 \\ + (a_{12} + a_{34})(x_1x_2 \text{ ou } x_3x_4) + (a_{13} + a_{24})(x_1x_3 \text{ ou } x_2x_4) + (a_{14} + a_{23})(x_1x_4 \text{ ou } x_2x_3)$$

La valeur numérique de ces regroupements est la somme des valeurs des coefficients. Par exemple, le contraste ℓ_1 est égal à l'effet du facteur 1 augmenté de la valeur de l'interaction 234 :

$$\ell_1 = a_1 + a_{234}$$

Si l'interaction 234 est faible, le contraste ℓ_1 est égal à l'effet du facteur 1. Comme nous avons admis l'hypothèse 1 qui suppose les interactions d'ordre 3 négligeables, on peut dire que le contraste ℓ_1 est égal à l'effet du facteur 1.

Considérons maintenant l'interaction 12 qui est aliée avec l'interaction 34 :

$$\ell_{12} = \ell_{34} = a_{12} + a_{34}$$

La valeur numérique du contraste ℓ_{12} (ou ℓ_{34}) est due à la fois à l'interaction 12 et à l'interaction 34. Il n'y a aucun moyen de savoir l'importance individuelle des interactions. Elles sont regroupées dans un même contraste et l'on ne connaît que la valeur de ce regroupement.

Mais à force de faire des plans d'expériences, les expérimentateurs ont constaté que, dans la plupart des cas, si un contraste avait une valeur numérique faible, c'est que les interactions aliées dans ce contraste étaient, elles-mêmes, faibles. Ils ont constaté également que lorsque deux facteurs étaient forts, il pouvait très bien y avoir une interaction entre eux. En revanche, si deux facteurs étaient faibles, il y avait peu de chance qu'il y ait une interaction entre eux. On en a déduit de nouvelles hypothèses d'interprétation des plans fractionnaires.

6.2.2 Nouvelles hypothèses d'interprétation

Pour analyser les résultats d'un plan fractionnaire, l'hypothèse 1 doit être complétée par de nouvelles hypothèses :

- **Hypothèse 1** – Les interactions d'ordre trois et toutes celles d'ordre plus élevé peuvent être négligées.
- **Hypothèse 2** – Tous les coefficients aliés dans un contraste faible (négligeable) sont eux-mêmes faibles (négligeables).
- **Hypothèse 3** – Si deux contrastes sont forts, on se méfiera de leur interaction qui peut être forte.
- **Hypothèse 4** – Si deux contrastes sont faibles, on suppose que leur interaction l'est aussi.
- **Hypothèse 5** – Un facteur principal faible et un facteur principal fort engendrent, le plus souvent mais pas toujours, une interaction faible.

Les cinq hypothèses présentées ici sont très souvent vérifiées mais il arrive parfois qu'elles soient mises en défaut. On estime que ces hypothèses sont valides à 95 % et qu'il y a environ 5 % de cas où elles ne sont pas vérifiées. Par exemple l'hypothèse 4 est parfois mise en défaut s'il faut que deux facteurs soient présents simultanément pour déclencher une variation de la réponse : la prise d'une résine thermodurcissable est déclenchée par la présence simultanée d'un catalyseur (premier facteur) et d'un accélérateur (deuxième facteur) ; le catalyseur seul ou l'accélérateur seul ne modifie pas la résine. Il est toujours possible d'adopter d'autres hypothèses en fonction du problème traité et des risques encourus. Pour une bonne analyse des résultats, il est prudent de toujours bien préciser les hypothèses d'interprétation que l'on a retenues. Ces hypothèses sont d'excellents guides pour analyser les résultats d'un plan fractionnaire. Elles permettent de trouver les insuffisances de l'expérimentation et de préciser les essais complémentaires à entreprendre pour obtenir une interprétation complète.

Nous venons de constater qu'un contraste est la somme de deux coefficients. Mais il peut aussi être une suite de sommes ou de différences car les associations entre coefficients dépendent de la position des points d'expériences qui ont été choisis lors de la construction du plan fractionnaire. Savoir retrouver les éléments qui composent les contrastes est primordial pour faire une bonne interprétation d'un plan fractionnaire. La théorie mathématique générale des alias fait appel à des notions qui dépassent le cadre de cet ouvrage. Il existe aussi une théorie beaucoup plus simple basée sur le calcul de Box. Elle convient uniquement aux plans factoriels fractionnaires à deux niveaux parfaitement fixés à -1 et $+1$ en valeurs codées. On peut aussi utiliser un logiciel de plan d'expériences qui donnent directement la composition des alias.

Les lecteurs intéressés par la théorie mathématique générale des alias se reporteront aux ouvrages signalés dans la bibliographie. Nous donnerons un aperçu du calcul de Box car, associé à la relation d'équivalence, il sert à comprendre l'interprétation et la construction des plans fractionnaires.

6.3 Calcul de Box

Le calcul de Box permet de retrouver rapidement comment les effets et les interactions sont aliés dans les contrastes. Comme nous l'avons déjà signalé, il s'applique aux plans fractionnaires à deux niveaux fixés à -1 et $+1$.

6.3.1 Notation de Box

Dans la notation de Box, on désigne par le chiffre **1** (en gras) la colonne des signes du facteur 1, signes ordonnés selon la présentation de Yates. Pour un plan 2^2 , on a :

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

La colonne du facteur 2 sera désignée par :

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

On introduit également des colonnes de signes moins et de signes plus :

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

6.3.2 Opérations sur les colonnes de signes

Le calcul de Box consiste à faire des opérations sur les colonnes de signes que nous venons de définir. On multiplie les signes terme à terme en appliquant la règle des signes. Par exemple, la multiplication de $\mathbf{1}$ par $\mathbf{2}$ est :

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

Le signe +1 du premier terme de la colonne-produit est le résultat de la multiplication de -1 (premier terme de $\mathbf{1}$) par -1 (premier terme de $\mathbf{2}$). Les autres signes de la colonne-produit sont obtenus de la même manière.

Voyons dans les paragraphes suivants les opérations les plus utilisées sur les colonnes de signes.

■ Multiplication d'une colonne par elle-même

Par exemple, multiplions la colonne $\mathbf{1}$ par elle-même :

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

On obtient une colonne de signes plus. La multiplication de la colonne $\mathbf{2}$ par elle-même donne le même résultat. D'une manière générale, la multiplication d'une colonne par elle-même donne une colonne de signes plus.

■ Multiplication d'une colonne par une colonne de signes plus

Multiplions la colonne **2** par **I** :

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

La multiplication d'une colonne par une colonne de signes plus donne la colonne initiale. On peut dire aussi que la multiplication par une colonne de signes plus ne change pas la colonne multipliée.

■ Multiplication d'une colonne par une colonne de signes moins

Multiplions la colonne **2** par **-I** :

$$\mathbf{2} \times (-\mathbf{I}) = \begin{bmatrix} -1 \\ -1 \\ +1 \\ +1 \end{bmatrix} \times \begin{bmatrix} -1 \\ -1 \\ -1 \\ -1 \end{bmatrix} = \begin{bmatrix} +1 \\ +1 \\ -1 \\ -1 \end{bmatrix} = -\mathbf{2}$$

La multiplication d'une colonne par une colonne de signes moins donne une colonne opposée à la colonne initiale.

■ Multiplication d'une colonne de signes plus par elle-même

On a :

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

La multiplication d'une colonne de signes plus par elle-même donne une colonne de signes plus.

■ Commutativité de la multiplication

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

Le résultat de la multiplication de **1** par **2** est le même que celui de la multiplication de **2** par **1**. Cette propriété s'appelle la *commutativité*.

6.3.3 Règles à retenir

Il faut simplement retenir les règles suivantes du calcul de Box. Pour simplifier, les signes multipliés ont été éliminés des formules :

- **Règle 1** – Commutativité :

$$1\ 2 = 2\ 1$$

- **Règle 2** – Multiplication d'une colonne par elle-même :

$$1\ 1 = I$$

$$2\ 2 = I$$

- **Règle 3** – Multiplication d'une colonne par **I** ou **-I** :

$$1\ I = 1$$

$$2\ I = 2$$

$$1\ (-I) = -1$$

$$2\ (-I) = -2$$

$$I\ I = I$$

6.4 Relation d'équivalence

6.4.1 Plan de base

Considérons un plan 2^3 et son modèle postulé :

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

Le terme x_1x_2 est le produit des niveaux des facteurs 1 et 2. Ce terme est celui de l'interaction 12. On peut donc construire une colonne de signes correspondant à l'interaction 12 en multipliant, selon le calcul de Box, les colonnes du facteur 1 et du facteur 2. On peut faire de même pour les interactions 13, 23 et 123. On peut aussi ajouter une colonne de signes plus pour introduire la constante dans les calculs. On obtient ainsi un tableau, appelé *matrice de base* ou *plan de base*, qui comporte 8 lignes et 8 colonnes (Tableau 6.4).

Tableau 6.4 – Matrice de base d'un plan 2^3 .

Essai n°	I	1	2	3	12	13	23	123
1	+	-	-	-	+	+	+	-
2	+	+	-	-	-	-	+	+
3	+	-	+	-	-	+	-	+
4	+	+	+	-	+	-	-	-
5	+	-	-	+	+	-	-	+
6	+	+	-	+	-	+	-	-
7	+	-	+	+	-	-	+	-
8	+	+	+	+	+	+	+	+

Cette matrice de base correspond à la matrice qui est utilisée pour calculer les effets et les interactions d'un plan complet. Mais pour un plan fractionnaire, on n'utilise que la moitié des essais de cette matrice. Divisons cette matrice de base en deux fractions : un demi-plan constitué des essais n° 5, 2, 3 et 8 (lignes plus foncées du tableau 6.5) et un demi-plan constitué des essais n° 1, 6, 7 et 4.

Tableau 6.5 – Matrice de base coupée en deux plans fractionnaires.

Essai n°	1	1	2	3	12	13	23	123
5	+	-	-	+	+	-	-	+
2	+	+	-	-	-	-	+	+
3	+	-	+	-	-	+	-	+
8	+	+	+	+	+	+	+	+
1	+	-	-	-	+	+	+	-
6	+	+	-	+	-	+	-	-
7	+	-	+	+	-	-	+	-
4	+	+	+	-	+	-	-	-

Dans un premier temps, intéressons-nous au demi-plan supérieur, le plan plus foncé. Il contient huit colonnes de quatre signes qui sont égales deux à deux. En notation de Box, on peut écrire, par exemple, que la colonne des quatre signes du facteur 1 est égale à celle de l'interaction 23, soit :

$$1 = 23$$

Nous avons vu que, dans ce demi-plan, le facteur 1 était confondu avec l'interaction 23 et que le contraste ℓ_1 calculé avec la colonne 1 (ou avec la colonne 23) était la somme des coefficients a_1 et a_{23} :

$$\ell_1 = \ell_{23} = a_1 + a_{23}$$

6.4.2 Relation d'équivalence

On remarque que les coefficients aliasés dans un contraste sont ceux qui ont les mêmes colonnes de signes dans le demi-plan considéré. C'est une règle générale : un contraste est la somme algébrique des coefficients qui ont les mêmes colonnes de signes. C'est-à-dire que si l'on a $1 = 23$ en calcul de Box, le contraste calculé avec la colonne 1 sera la somme $\ell_1 = a_1 + a_{23}$ et le contraste calculé avec la colonne 23 sera la somme $\ell_{23} = a_1 + a_{23}$. On a donc :

$$1 = 23 \text{ est équivalent à } \ell_1 = \ell_{23} = a_1 + a_{23}$$

C'est la relation d'équivalence. Elle est valable dans les deux sens et elle constitue la base de la théorie des aliasés.

L'examen du tableau 6.5 montre que l'on a aussi :

$$2 = 13 \text{ est équivalent à } \ell_2 = \ell_{13} = a_2 + a_{13}$$

$$3 = 12 \text{ est équivalent à } \ell_3 = \ell_{12} = a_3 + a_{12}$$

$$I = 123 \text{ est équivalent à } \ell_0 = \ell_{123} = a_0 + a_{123}$$

L'importance de ces relations réside dans le fait que l'on peut retrouver comment les effets et les interactions sont aliasés dans chaque contraste. Pour connaître la structure des aliasés, on pourrait opérer comme nous venons de le faire : écrire un plan de base, le découper en fractions, rechercher les colonnes de signes qui sont égales. Mais ce processus utile pour comprendre la relation d'équivalence est à la fois trop long et trop laborieux pour retrouver la structure des aliasés. C'est avec l'utilisation du calcul de Box et l'introduction des générateurs d'aliasés que nous allons facilement retrouver cette structure.

Remarque

Le terme *contraste* n'est pas universellement utilisé. Dans certains ouvrages, on trouvera le terme *confusion*. On trouvera également le terme *aliasé* très souvent utilisé. Enfin, le terme *concomitance* est parfois employé pour montrer qu'on estime simultanément des coefficients dans la même structure. Tous ces termes sont synonymes et traduisent les deux aspects de la relation d'équivalence en soulignant que des effets et des interactions sont fusionnés dans une seule inconnue et que des colonnes de signes se correspondent dans les demi-plans. La notion de contrastes, elle-même, est délicate et les différents termes utilisés expriment la difficulté à trouver le bon terme. Ces différentes appellations sont utilisées et considérées comme des synonymes dans le présent ouvrage.

6.5 Générateurs d'aliasés

6.5.1 Générateur d'aliasés du demi-plan supérieur

Considérons le demi-plan supérieur du tableau 6.5. Les deux colonnes de signes plus permettent d'écrire en notation de Box :

$$I = 123 \tag{6.3}$$

En multipliant par **1** les deux membres de cette relation et en tenant compte des règles de multiplication de Box, on retrouve les colonnes identiques du demi-plan supérieur :

1. Multiplication par **1** des deux membres de la relation {6.3} :

$$1.I = 1.123$$

2. Application de la règle 2 :

$$1.I = I.23$$

3. Application de la règle 3 :

$$1 = 23$$

En multipliant les deux membres de la relation {6.3} successivement par **2** et par **3**, on obtient les deux relations :

$$2 = 13$$

$$3 = 12$$

La relation $\mathbf{I} = \mathbf{123}$ permet donc de retrouver toutes les colonnes qui sont égales dans le demi-plan supérieur. Connaissant ces égalités et en appliquant la relation d'équivalence, on retrouve comment les coefficients sont aliés dans les contrastes. La relation $\mathbf{I} = \mathbf{123}$ s'appelle le *générateur d'aliés*.

6.5.2 Générateur d'aliés du demi-plan inférieur

En considérant le demi-plan inférieur (Tableau 6.5), on vérifie que les colonnes se correspondent deux à deux, mais avec des signes opposés. À la colonne de signes plus correspond la colonne de signes moins de l'interaction 123. On peut écrire, en notation de Box :

$$\mathbf{I} = -\mathbf{123} \quad \{6.4\}$$

C'est le générateur d'aliés du demi-plan inférieur.

En multipliant par $\mathbf{1}$ les deux membres de la relation {6.4} et en tenant compte des règles de multiplication de Box, on a :

$$\mathbf{1.I} = \mathbf{1.-123}$$

$$\mathbf{1} = -\mathbf{I} \mathbf{23}$$

$$\mathbf{1} = -\mathbf{23}$$

La colonne de l'interaction 23 a bien des signes opposés à ceux de la colonne 1. On trouve la structure du contraste de la colonne 1 en appliquant la relation d'équivalence :

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

Les contrastes calculés avec le demi-plan inférieur sont des différences de coefficients. On calcule que l'on a de même :

$$\mathbf{2} = -\mathbf{13} \text{ soit } \ell'_2 = -\ell'_{13} = a_2 - a_{13}$$

$$\mathbf{3} = -\mathbf{12} \text{ soit } \ell'_3 = -\ell'_{12} = a_3 - a_{12}$$

$$\mathbf{I} = -\mathbf{123} \text{ soit } \ell'_0 = -\ell'_{123} = a_0 - a_{123}$$

6.5.3 Lecture des aliés dans les logiciels

Les bons logiciels de plans d'expériences indiquent toujours comment les coefficients sont aliés dans les contrastes. Par exemple, le logiciel joint à ce livre donne les aliés, en même temps que les résultats du plan, dans une rubrique intitulée *Singularity Details*. Les colonnes qui sont égales sont indiquées comme dans le calcul de Box. D'autres logiciels fournissent les coefficients aliés dans chaque contraste.

6.6 Construction pratique d'un plan fractionnaire

Lorsqu'on examine les quatre premières colonnes du demi-plan supérieur (Tableau 6.5), on constate que l'on retrouve les colonnes de signes d'un plan de base 2^2 . Ce plan comporte la colonne des facteurs 1 et 2, la colonne 12 construite par la règle des signes et une colonne de signes plus (Tableau 6.6).

Tableau 6.6 – Plan de base 2^2 .

Essai n°	I	1	2	12
1	+	-	-	+
2	+	+	-	-
3	+	-	+	-
4	+	+	+	+

L'interaction **12** peut servir à étudier un facteur supplémentaire, le facteur 3. En effet, les niveaux d'étude du facteur supplémentaire sont semblables aux signes de la colonne de signes de l'interaction 12 (Tableau 6.7). On profite de cette similitude pour construire facilement les plans fractionnaires. En notation de Box, on écrit que le facteur 3 est étudié sur les signes de l'interaction 12 :

$$\mathbf{3} = \mathbf{12}$$

Multiplions chaque membre de cette relation par **3**, on retrouve le générateur du plan fractionnaire :

$$\mathbf{I} = \mathbf{123}$$

Ce générateur permet, grâce au calcul de Box et à la relation d'équivalence, de retrouver comment les coefficients sont aliésés dans les contrastes.

Tableau 6.7 – Plan d'expériences 2^{3-1} bâti avec le générateur $\mathbf{I} = \mathbf{123}$.

Essai n°	1	2	3
1	-	-	+
2	+	-	-
3	-	+	-
4	+	+	+

Au lieu de prendre le générateur $\mathbf{I} = \mathbf{123}$, on peut prendre le générateur :

$$\mathbf{I} = -\mathbf{123}$$

Dans ce cas, le troisième facteur est étudié sur les signes de l'interaction -12 (Tableau 6.8).

Tableau 6.8 – Plan d'expériences 2^{3-1} bâti avec le générateur $\mathbf{I} = -\mathbf{123}$.

Essai n°	1	2	3
1	-	-	-
2	+	-	+
3	-	+	+
4	+	+	-

La construction pratique des plans fractionnaires que nous venons de voir sur un plan de base 2^2 , se généralise à tous les plans de base 2^k :

- On choisit un plan complet et l'on écrit le plan de base correspondant en appliquant la règle des signes.
- Dans ce plan de base, on choisit une colonne de signes correspondant à une interaction et on lui attribue un facteur supplémentaire. On peut prendre, soit la colonne elle-même avec ses signes, soit la colonne avec les signes opposés.
- On en déduit le générateur d'aliases grâce au calcul de Box.
- En multipliant le générateur d'aliases et en appliquant la relation d'équivalence, on retrouve la structure de tous les contrastes.

Voyons comment cela fonctionne sur un plan de base 2^3 .

Reprenons le plan de base 2^3 (Tableau 6.9). Il y a quatre colonnes disponibles pour étudier des facteurs supplémentaires : les interactions **12**, **13**, **23** et **123**. On peut donc construire plusieurs plans fractionnaires qui auront tous 8 essais mais qui permettront d'étudier plus ou moins de facteurs.

Tableau 6.9 – Plan de base d'un plan 2^3 .

Essai n°	I	1	2	3	12	13	23	123
1	+	-	-	-	+	+	+	-
2	+	+	-	-	-	-	+	+
3	+	-	+	-	-	+	-	+
4	+	+	+	-	+	-	-	-
5	+	-	-	+	+	-	-	+
6	+	+	-	+	-	+	-	-
7	+	-	+	+	-	-	+	-
8	+	+	+	+	+	+	+	+

De ce plan de base, nous allons successivement extraire le plan complet et les plans fractionnaires que l'on peut construire.

6.6.1 Plan complet 2^3

Les colonnes 1, 2 et 3 sont celles d'un plan complet pour étudier trois facteurs (Tableau 6.10).

Tableau 6.10 – Plan complet 2^3 .

Essai n°	1	2	3
1	-	-	-
2	+	-	-
3	-	+	-
4	+	+	-
5	-	-	+
6	+	-	+
7	-	+	+
8	+	+	+

6.6.2 Plan fractionnaire 2^{4-1}

C'est le plan qui a été utilisé pour le tellure. On désire étudier quatre facteurs en ne faisant que 8 essais. On donne le nom de 4 au facteur supplémentaire. On choisit l'une des quatre colonnes d'interaction, par exemple la colonne **123** (Tableau 6.11). Les niveaux d'étude du facteur 4 sont ceux de la colonne **123**. On construit le plan fractionnaire en extrayant les quatre colonnes **1**, **2**, **3** et **123** du plan de base.

Tableau 6.11 – Plan d'expériences 2^{4-1} .

Essai n°	1	2	3	4 = 123
1	-	-	-	-
2	+	-	-	+
3	-	+	-	+
4	+	+	-	-
5	-	-	+	+
6	+	-	+	-
7	-	+	+	-
8	+	+	+	+

Le quatrième facteur étant étudié sur la colonne de signes de l'interaction 123, on peut écrire en notation de Box :

$$\mathbf{I} = 1234$$

C'est le générateur d'alias qui permet de calculer la structure des contrastes. En multipliant ce générateur successivement par **1**, **2**, **3** et **4**, on obtient les égalités de colonnes, puis en utilisant la relation d'équivalence, les contrastes :

1 = 234	est équivalent à	$\ell_1 = a_1 + a_{234}$
2 = 134	est équivalent à	$\ell_2 = a_2 + a_{134}$
3 = 124	est équivalent à	$\ell_3 = a_3 + a_{124}$
4 = 123	est équivalent à	$\ell_4 = a_4 + a_{123}$
I = 1234	est équivalent à	$\ell_0 = a_0 + a_{1234}$

On aurait pu aliaser le facteur 4 sur une autre interaction, on aurait eu d'autres structures de contrastes. Par exemple, si on choisit d'étudier le facteur 4 sur l'interaction 12, on a :

$$\mathbf{4} = 12$$

d'où le générateur d'alias :

$$\mathbf{I} = 124$$

En multipliant ce générateur successivement par **1**, **2**, **3** et **4**, on obtient les relations d'équivalence et la structure des contrastes correspondants :

1 = 24	est équivalent à	$\ell_1 = a_1 + a_{24}$
2 = 14	est équivalent à	$\ell_2 = a_2 + a_{14}$
3 = 1234	est équivalent à	$\ell_3 = a_3 + a_{1234}$
4 = 12	est équivalent à	$\ell_4 = a_4 + a_{12}$
I = 124	est équivalent à	$\ell_0 = a_0 + a_{124}$

On pourrait également choisir soit l'interaction 13 soit l'interaction 23 pour étudier le facteur 4. Mais ces choix ne sont pas équivalents. Le premier choix, **4 = 123**, est bien meilleur que les trois autres car les effets principaux des facteurs sont tous aliésés avec des interactions d'ordre 3, c'est-à-dire avec des interactions qui ont de grandes chances d'être faibles. Ce n'est pas le cas des trois autres plans 2^{4-1} pour lesquels certains effets principaux sont aliésés avec des interactions d'ordre 2.

6.6.3 Plan fractionnaire 2^{5-2}

Il est tout à fait possible d'étudier deux facteurs supplémentaires. On choisit deux colonnes qui ne sont pas déjà occupées par des facteurs (ici colonnes 1, 2 et 3), par exemple, on peut choisir la colonne **12** pour le quatrième facteur et la colonne **13** pour le cinquième facteur (Tableau 6.12). Le plan fractionnaire correspondant est un plan 2^{5-2} car il s'agit d'un plan complet à 32 essais coupé en quatre. Le plan 2^5 a été divisé par 2^2 soit $2^5/2^2$, ce qui peut également s'écrire 2^{5-2} . Le tableau 6.12 représente ce plan fractionnaire 2^{5-2} .

Tableau 6.12 – Plan d'expériences 2^{5-2} .

Essai n°	1	2	3	4 = 12	5 = 13
1	-	-	-	+	+
2	+	-	-	-	-
3	-	+	-	-	+
4	+	+	-	+	-
5	-	-	+	+	-
6	+	-	+	-	+
7	-	+	+	-	-
8	+	+	+	+	+

Établissons, grâce à la notation de Box, la structure des aliases de ce plan fractionnaire. On a deux relations :

$$4 = 12 \qquad 5 = 13$$

et, bien sûr, deux générateurs d'aliases :

$$I = 124 \qquad I = 135$$

Ces deux générateurs d'aliases proviennent du choix de deux interactions indépendantes du plan de base. On les appelle des générateurs d'aliases *indépendants*.

Si l'on multiplie ces deux générateurs d'aliases indépendants membre à membre, on obtient un troisième générateur :

$$I.I = 124.135 \\ I = 2345$$

Ce nouveau générateur d'aliases est un générateur *dépendant*. On remarque que deux facteurs supplémentaires introduisent non pas un générateur mais un *groupe de générateurs d'aliases* (GGA) comportant quatre termes :

$$I = 124 = 135 = 2345$$

Ce GGA est utilisé pour établir la structure des aliases de ce plan fractionnaire. Par exemple, le contraste ℓ_1 sera déterminé en multipliant tous les termes du GGA par la colonne **I** :

$$1.I = 1.124 = 1.135 = 1.2345$$

En simplifiant :

$$1 = 24 = 35 = 12345$$

Les quatre colonnes **1**, **24**, **35** et **12345** sont identiques dans le quart de plan de base. La relation d'équivalence donne les quatre coefficients aliésés dans les contrastes ℓ_1 , ℓ_{24} , ℓ_{35} et ℓ_{12345} :

$$\ell_1 = \ell_{24} = \ell_{35} = \ell_{12345} = a_1 + a_{24} + a_{35} + a_{12345}$$

On simplifie l'écriture en n'indiquant que l'un des contrastes :

$$\ell_1 = a_1 + a_{24} + a_{35} + a_{12345}$$

Les autres contrastes sont calculés de la même manière et comportent tous 4 coefficients.

Remarque

Le *groupe de générateurs d'alias* (GGA) est également appelé *relation de définition*. En effet, c'est cette relation qui permet de définir le plan fractionnaire avec précision et qui donne la structure des alias.

6.6.4 Plan fractionnaire 2^{7-4}

On peut généraliser cette méthode et utiliser toutes les colonnes du plan de base (Tableau 6.13). Sur le plan de base 2^3 , on peut étudier jusqu'à 7 facteurs puisqu'il y a 4 colonnes d'interaction. Dans ce cas, il s'agit de diviser en 16 un plan complet de 128 essais.

Tableau 6.13 – Plan d'expériences 2^{7-4} .

Essai n°	1	2	3	4 = 123	5 = 12	6 = 13	7 = 23
1	-	-	-	-	+	+	+
2	+	-	-	+	-	-	+
3	-	+	-	+	-	+	-
4	+	+	-	-	+	-	-
5	-	-	+	+	+	-	-
6	+	-	+	-	-	+	-
7	-	+	+	-	-	-	+
8	+	+	+	+	+	+	+

Le générateur d'alias s'obtient en prenant d'abord les 4 générateurs indépendants :

$$4 = 123$$

$$5 = 12$$

$$6 = 13$$

$$7 = 23$$

Les générateurs dépendants s'obtiennent en multipliant deux à deux, puis trois à trois et enfin quatre à quatre les générateurs indépendants. Le GGA contient 16 termes. Ce GGA est multiplié successivement par **1**, **2**, **3**, **4**, **5**, **6** et **7**. Les relations obtenues permettent d'établir la structure des contrastes qui comportent chacun 16 termes dans le cas du plan 2^{7-4} . Dans le cas général, les contrastes d'un plan fractionnaire 2^{k-p} contiennent 2^p termes. Le calcul de tels contrastes est ennuyeux, fatigant et fastidieux et nous verrons comment on les obtient facilement avec le logiciel.

6.7 Nombre maximal de facteurs étudiés sur un plan de base

On peut étudier autant de facteurs supplémentaires qu'il y a d'interactions dans le plan de base (Tableau 6.14).

Sur un plan de base 2^2 il y a une interaction. On pourra donc étudier trois facteurs, deux sur les colonnes 1 et 2, le troisième sur la colonne de l'interaction.

Sur un plan de base 2^3 , il y a quatre interactions. On pourra donc étudier sept facteurs, trois sur les colonnes 1, 2 et 3, les quatre autres sur les colonnes d'interaction 12, 13, 23 et 123.

Le tableau 6.14 indique le nombre maximal de facteurs que l'on peut étudier sur différents plans de base.

Tableau 6.14 – Nombre maximal de facteurs étudiés sur un plan de base donné.

Plan de base	Nombre de facteurs du plan complet	Nombre d'interactions	Nombre maximal de facteurs étudiés
2^2	2	1	3
2^3	3	4	7
2^4	4	11	15
2^5	5	26	31
2^6	6	57	63
2^7	7	120	127

6.8 Application de la théorie des aliases à l'exemple 05

Voyons comment toutes les notions acquises sont mises en œuvre sur l'exemple de ce chapitre.

6.8.1 Confusions du plan Tellure

Le plan Tellure est un plan 2^{4-1} dont le générateur d'aliases est :

$$I = 1234$$

En appliquant le calcul de Box, on obtient pour les facteurs principaux :

$$1 = 234 \qquad 2 = 134 \qquad 3 = 124$$

Si l'on a choisi un modèle mathématique sans les interactions d'ordre 3, on a immédiatement (relation d'équivalence) :

$$\ell_1 = a_1 \qquad \ell_2 = a_2 \qquad \text{et } \ell_3 = a_3$$

Pour les interactions, on obtient :

$$12 = 34 \qquad 13 = 24 \qquad 23 = 14$$

d'où :

$$\ell_{12} = \ell_{34} = a_{12} + a_{34} \qquad \ell_{13} = \ell_{24} = a_{13} + a_{24} \qquad \text{et } \ell_{23} = \ell_{14} = a_{23} + a_{14}$$

6.8.2 Application des hypothèses d'interprétation au plan Tellure

L'hypothèse 1 a été appliquée dans le choix du modèle. On obtient directement les effets des facteurs principaux. Ils ne sont pas confondus avec des interactions. Pour faire l'analyse des résultats reprenons le tableau 6.3 des contrastes sous forme d'un diagramme à barres (Figure 6.1) et faisons en l'analyse.

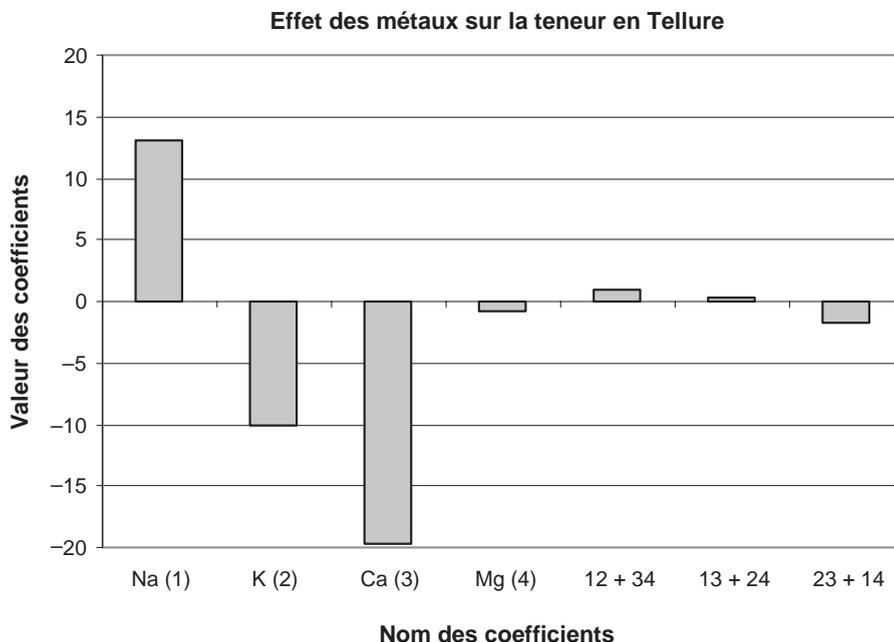


Figure 6.1 – Diagramme à barres des coefficients.

Appliquons l'hypothèse 2 (Tous les coefficients aliasés dans un contraste faible sont eux-mêmes faibles) en admettant qu'un contraste inférieur à 2 est négligeable.

Le contraste ℓ_{13} étant égal à 0,25, on le considère comme négligeable. Par conséquent, les coefficients a_{13} et a_{24} sont également négligeables. On tient le même raisonnement pour le contraste ℓ_{14} et on en déduit que les coefficients a_{14} et a_{23} sont proches de zéro et donc négligeables. De la même façon, les coefficients a_{12} et a_{34} sont négligeables.

6.9 Conclusion de l'étude

Seuls les sels de calcium, de potassium et de sodium faussent les mesures de la concentration en tellure de l'eau de mer. Il n'y a pas d'interaction entre ces métaux. Tout se passe comme si chaque métal était seul à modifier la mesure du tellure dans l'eau de mer :

- Le magnésium ne modifie pas les mesures et on obtient des valeurs correctes de la teneur en tellure.

- Le facteur le plus influent est le calcium. La teneur apparente en tellure est plus faible que la teneur réelle. En présence de calcium, on obtient des valeurs par défaut.
- Le potassium a la même influence que le calcium. En présence de potassium, on obtient des valeurs par défaut.
- Quant au sodium il conduit à une teneur apparente en tellure supérieure à la valeur réelle. Il induit des valeurs trop fortes. En présence de sodium, on obtient des valeurs par excès.

Pour faire des mesures précises on ne peut pas négliger ces interférences. Les sels de calcium et de potassium diminuent les valeurs réelles. Au contraire, les sels de sodium augmentent les valeurs réelles.

L'étude d'orientation qui vient d'être faite montre qu'il faut envisager des corrections aux résultats de mesure. Il faut entreprendre une étude plus complète pour pouvoir donner des directives précises de corrections. Il faut alors envisager un ou plusieurs plans permettant de modéliser le phénomène. Nous verrons que les plans composites sont alors précieux pour établir un modèle du second degré (chapitre 9).

7 • EXEMPLES DE PLANS FACTORIELS FRACTIONNAIRES

Dans le chapitre précédent, nous avons vu un exemple de plan fractionnaire (le plan Tellure). Nous avons également vu plusieurs aspects de la théorie concernant ces plans : la théorie des alias, le calcul de Box, la relation d'équivalence, les hypothèses d'interprétation et la construction des plans fractionnaires. L'assimilation de toutes ces notions ne peut pas se faire avec un seul exemple. C'est pourquoi le présent chapitre traite plusieurs exemples de plans fractionnaires. Le nombre de facteurs de ces plans ira croissant sans que pour autant le nombre d'essais soit trop élevé. L'interprétation de ces plans sera conduite en tenant compte d'un modèle postulé, de la théorie des alias et des hypothèses d'interprétation. Mais nous verrons qu'avec très peu d'essais on peut obtenir l'information cherchée.

Nous avons vu que sur un plan de base 2^{k-p} on pouvait théoriquement étudier k facteurs dont p avaient pour niveaux d'étude les signes des colonnes de certaines interactions. Les signes moins sont utilisés comme niveaux bas et les signes plus comme niveaux hauts. Jusqu'à maintenant nous n'avons utilisé qu'une seule interaction du plan de base. Sur un plan de base 2^3 il y a 4 interactions, on peut donc étudier 4 facteurs supplémentaires, soit 7 facteurs en 8 essais. Le plan fractionnaire correspondant est un plan 2^{7-4} . Comme toutes les interactions sont utilisées pour étudier des facteurs, on dit que le plan est *saturé*. Est-il vraiment possible d'étudier 7 facteurs en seulement 8 essais ? Nous allons voir que cela est possible et ajoutons que, dans la pratique, les expérimentateurs réalisent souvent de tels plans. La seule condition pour réussir à tous les coups, est de bien utiliser la théorie des alias et de bien manier les hypothèses d'interprétation. Il arrive en effet souvent que l'on soit obligé de poursuivre la première expérimentation par un plan complémentaire bien choisi pour lever les ambiguïtés qui apparaissent lors de l'interprétation du plan initial. Mais, même si l'on réalise un premier plan de 8 essais suivi d'un plan complémentaire également de 8 essais, on est toujours gagnant car, en définitive, on exécute 16 essais au lieu des 128 d'un plan complet 2^7 . On retrouve le pari toujours gagnant des plans fractionnaires.

7.1 Exemple 06 : la sulfonation

7.1.1 Préparation du plan d'expériences

■ Présentation de l'étude : sulfonation d'un mélange BTX

La sulfonation des composés aromatiques permet d'obtenir des composés tensioactifs qui sont les produits de base des lessives du commerce. Les composés aromatiques étudiés sont le benzène, le toluène et les xylènes (mélange BTX) sous-produits d'une exploitation de pétrole brut. Seuls le toluène et les xylènes fournissent des tensioactifs intéressants. Pour obtenir les produits tensioactifs, on ajoute de l'acide sulfurique (SO_4H_2) au mélange BTX, puis on laisse ces produits en contact pendant un certain temps pour qu'ils réagissent. En présence d'acide sulfurique, les trois produits benzène, toluène et xylènes se transforment progressivement en acides sulfoniques qui serviront à préparer les sulfonates. La réaction de sulfonation peut être accélérée soit par ajout d'anhydride sulfurique (SO_3), soit par élimination de l'eau formée, soit simultanément par ces deux traitements. La réaction chimique des BTX avec l'acide sulfurique s'appelle la *sulfonation* et on dit que le mélange BTX a été *sulfoné*. Les sulfonates formés contiennent les tensioactifs.

■ Définition de l'objectif de l'étude

On cherche à obtenir les sulfonates de toluène et de xylènes qui permettent de préparer des lessives de bonne qualité. En revanche, on cherche à éviter les sulfonates de benzène qui ne présentent aucun intérêt. On recherche donc les conditions opératoires qui permettent de sulfoner le toluène et les xylènes tout en évitant la sulfonation du benzène.

L'étude sera réussie si on peut donner des consignes de réglage des facteurs pour atteindre les deux objectifs : sulfoner le toluène et les xylènes ; ne pas sulfoner le benzène.

■ Choix de la réponse permettant d'atteindre l'objectif

La réponse choisie est le rapport de sulfonation défini par :

$$\rho = \frac{(\text{Toluène} + \text{xylènes}) \text{ sulfonés}}{\text{Benzène sulfoné}}$$

On a donc tout intérêt à obtenir une grande valeur de ce rapport.

■ Recherche des facteurs qui pourraient être influents sur la réponse

Parmi tous les facteurs pouvant influencer la réaction de sulfonation, l'expérimentateur en a retenu six :

- la durée de la mise en contact de l'acide sulfurique avec le mélange BTX,
- le pourcentage de SO_3 dans l'acide sulfurique,
- la température de la réaction,
- l'élimination de l'eau en cours de réaction,
- la durée d'addition d'acide sulfurique,
- le rapport (acide)/(toluène + xylènes).

Le domaine d'étude des six facteurs étudiés au cours du plan d'expériences est précisé dans le tableau 7.1.

Tableau 7.1 – Facteurs retenus et domaine d'étude.

Facteurs	Niveau -1	Niveau +1
Durée de la réaction (1)	5 h	9 h
SO ₃ dans SO ₄ H ₂ (2)	0 %	10 %
Température de la réaction (3)	95 °C	110 °C
Élimination de l'eau (4)	sans	avec
Durée de l'addition d'acide (5)	10 min	30 min
Rapport (acide/toluène +xylènes) (6)	0,8	1,2

■ Choix du plan d'expériences

Si l'expérimentateur avait décidé de faire un plan complet pour des facteurs ne prenant que deux niveaux, il aurait dû réaliser 64 essais. Ce qui est considérable. Il a donc été décidé de réduire le nombre d'essais et de faire un plan factoriel fractionnaire. Les facteurs auront toujours deux niveaux d'étude. Le budget de l'expérimentateur autorise une étude de 16 essais, soit quatre fois moins d'essais que le plan complet. Avec un plan fractionnaire 2^{6-2} de 16 essais, il est tout à fait possible d'étudier 6 facteurs avec leurs interactions d'ordre 2.

■ Construction du plan d'expériences

Le plan de base est un plan 2^4 . L'expérimentateur choisit d'aliaser le facteur 5 sur l'interaction 124 et le facteur 6 sur l'interaction 234. C'est-à-dire que les niveaux d'étude du facteur 5 seront analogues aux signes de l'interaction 124. De même les niveaux d'étude du facteur 6 seront analogues aux signes de l'interaction 234. On peut écrire en notation de Box :

$$5 = 124$$

$$6 = 234$$

On écrit les 4 colonnes d'un plan 2^4 et on calcule les signes des interactions 124 et 234 en utilisant la règle des signes (Tableau 7.2).

7.1.2 Expérimentation

Voir tableau 7.2, page suivante.

Tableau 7.2 – Essais et réponses du plan d'expériences sur la sulfonation.

Essai n°	Durée 1	SO ₃ 2	Température 3	Eau 4	Acide 5 = 124	Rapport 6 = 234	Réponse
1	-1	-1	-1	-1	-1	-1	11,76
2	+1	-1	-1	-1	+1	-1	12,55
3	-1	+1	-1	-1	+1	+1	7,89
4	+1	+1	-1	-1	-1	+1	8,20
5	-1	-1	+1	-1	-1	+1	12,35
6	+1	-1	+1	-1	+1	+1	13,98
7	-1	+1	+1	-1	+1	-1	7,25
8	+1	+1	+1	-1	-1	-1	10,10
9	-1	-1	-1	+1	+1	+1	13,06
10	+1	-1	-1	+1	-1	+1	13,40
11	-1	+1	-1	+1	-1	-1	12,10
12	+1	+1	-1	+1	+1	-1	8,91
13	-1	-1	+1	+1	+1	-1	13,65
14	+1	-1	+1	+1	-1	-1	16,10
15	-1	+1	+1	+1	-1	+1	11,09
16	+1	+1	+1	+1	+1	+1	9,86
-1	5	0	95	Sans élim.	10	0,8	
+1	9	10	110	Avec élim.	30	1,2	

7.1.3 Interprétation des résultats des calculs

Pour interpréter les résultats d'un plan fractionnaire il est bon de connaître la manière dont les coefficients sont aliasés dans les contrastes. Le plus souvent, il ne s'agit que des interactions d'ordre 2 car l'hypothèse d'interprétation 1 a fait disparaître les interactions d'ordre 3.

■ Modèle mathématique

On adopte un modèle comprenant le coefficient constant, les coefficients des facteurs principaux et les coefficients des interactions d'ordre 2 :

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5 + a_6x_6 + a_{12}x_1x_2 + a_{13}x_1x_3 + \dots + a_{56}x_5x_6 \quad \{7.1\}$$

On a donc un système de 16 équations pour trouver 22 inconnues. On sait qu'il y a des confusions entre les interactions d'ordre 2 et le calcul de Box nous a indiqué comment ces interactions se regroupaient dans les aliasés.

■ Confusions

Nous avons déjà écrit :

$$5 = 124$$

$$6 = 234$$

d'où :

$$I = 1245$$

$$I = 2346$$

Si on multiplie **1245** par **2346**, on obtient **1356** qui est aussi égal à **I**. On obtient ainsi le groupe des générateurs d'aliasés :

$$I = 1245 = 2346 = 1356$$

Ce groupe montre que les effets des facteurs principaux sont aliasés avec des interactions d'ordre 3. Si l'hypothèse 1 est respectée, on obtiendra directement l'effet des facteurs principaux. Les interactions d'ordre 2 seront aliasées entre elles. On obtient les contrastes en multipliant le générateur d'aliasés par **12**, **13**... et en appliquant la relation d'équivalence. Les 15 interactions d'ordre 2 se regroupent dans les 7 contrastes suivants :

$$\ell_{12} = \ell_{45} = a_{12} + a_{45}$$

$$\ell_{13} = \ell_{56} = a_{13} + a_{56}$$

$$\ell_{14} = \ell_{25} = a_{14} + a_{25}$$

$$\ell_{15} = \ell_{24} = \ell_{36} = a_{15} + a_{24} + a_{36}$$

$$\ell_{16} = \ell_{35} = a_{16} + a_{35}$$

$$\ell_{23} = \ell_{46} = a_{23} + a_{46}$$

$$\ell_{26} = \ell_{34} = a_{26} + a_{34}$$

■ Valeur des effets

Le tableau 7.3 indique les résultats du calcul des effets et des alias. Le diagramme à barres de la figure 7.1 illustre ce tableau.

Tableau 7.3 – Effets et alias (unités codées).

Effet	Valeur
Constante	11,39
Durée de la réaction (1)	0,25
SO ₃ dans SO ₄ H ₂ (2)	-1,96
Température de la réaction (3)	0,41
Élimination de l'eau (4)	0,88
Durée de l'addition d'acide (5)	-0,49
Rapport acide/toluène +xylènes (6)	-0,16
$\ell_{12} = a_{12} + a_{45}$	-0,40
$\ell_{13} = a_{13} + a_{56}$	0,46
$\ell_{14} = a_{14} + a_{25}$	-0,45
$\ell_{15} = a_{15} + a_{24} + a_{36}$	0,18
$\ell_{16} = a_{16} + a_{35}$	-0,11
$\ell_{23} = a_{23} + a_{46}$	-0,26
$\ell_{26} = a_{26} + a_{34}$	-0,003

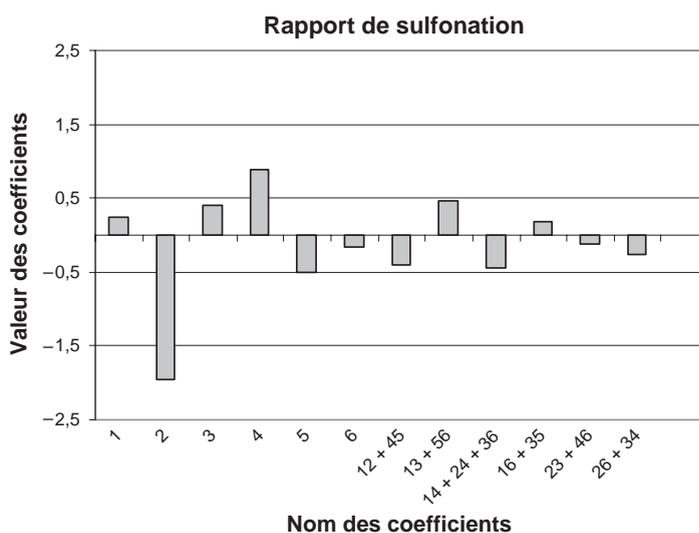


Figure 7.1 – Diagramme à barres des coefficients du modèle Sulfonation.

■ Comment distinguer les facteurs influents

Deux facteurs, l'anhydride sulfurique (facteur 2) et l'élimination de l'eau (facteur 4), sont plus élevés que les autres. Ce sont donc eux qui vont être les plus efficaces pour approcher ou atteindre l'objectif. Mais y a-t-il d'autres facteurs influents ? Savoir où mettre la limite entre les facteurs influents et les non influents est une question où délicate qui a été traitée lorsqu'on possède une évaluation de l'erreur expérimentale (chapitre 5). Comme on limite le nombre d'essais lors de l'exécution d'un plan fractionnaire, on ne possède pas toujours cette valeur. L'analyse de la variance apporte un début de réponse mais il ne faut pas oublier que les résidus, lorsqu'il y en a, sont constitués des interactions qui n'ont pas été calculées. En fonction du modèle postulé choisi, on peut obtenir une RMSE (racine carrée du carré moyen des résidus) plus ou moins élevée. L'analyse de la variance est une réponse partielle au problème posé. Il reste deux outils permettant de se faire une idée de l'importance relative des coefficients. Ce sont le diagramme de Pareto et le diagramme de Daniel qui vont nous aider à prendre une décision raisonnable. Ces outils aident l'expérimentateur à fixer une limite entre les facteurs influents et peut-être faiblement influents. Dans cette situation on peut tenir plusieurs types de raisonnement qui se complètent les uns les autres sans apporter chacun une réponse définitive.

Un premier raisonnement consiste à dire que tous les effets inférieurs à une certaine valeur ne sont pas assez élevés pour modifier la réponse de manière intéressante. Par exemple, le responsable dira qu'une variation du rapport (acide/toluène + xylènes) inférieure à 0,5 ne l'intéresse pas. Dans ce cas il n'y a que deux facteurs influents. Si le responsable décide de mettre la limite à 0,25 il y aura 4 facteurs influents.

La seconde méthode consiste à utiliser le diagramme de Pareto (Figure 7.2) : on ordonne les valeurs absolues des coefficients par ordre décroissant. On a donc une vue d'ensemble de tous les coefficients. Cela facilite le choix de la limite entre les deux populations, on conserve les coefficients qui sont au-dessus d'une certaine valeur et on élimine ceux qui sont au-dessous. Il n'y a pas une grande différence de fond entre ces deux premiers raisonnements à part que le second s'appuie sur un diagramme et qu'il facilite les comparaisons.

Une troisième méthode vient en appui des deux raisonnements précédents. On utilise un diagramme de Daniel : les coefficients de faibles valeurs suivent une distribution normale, ils sont assimilés aux erreurs expérimentales, et ils s'alignent sur la droite de Henry. Les coefficients de valeur élevée ne suivent pas une distribution normale et ils ne s'alignent pas sur la droite de Henry précédente. On distingue donc deux populations, les erreurs expérimentales et les effets à prendre en compte (Figure 7.3). Avec le diagramme de Daniel, l'arbitraire de la décision est considérablement diminué.

Si on en a la possibilité, il faut aussi employer les autres techniques statistiques qui aident à prendre la bonne décision. Par exemple, si l'expérimentateur a fait des répétitions, il a une estimation de l'erreur expérimentale et il peut utiliser cette donnée pour obtenir la limite entre les facteurs significatifs et ceux qui ne le sont pas. Il peut choisir pour limite, comme il est souvent fait, deux ou trois fois l'écart-type.

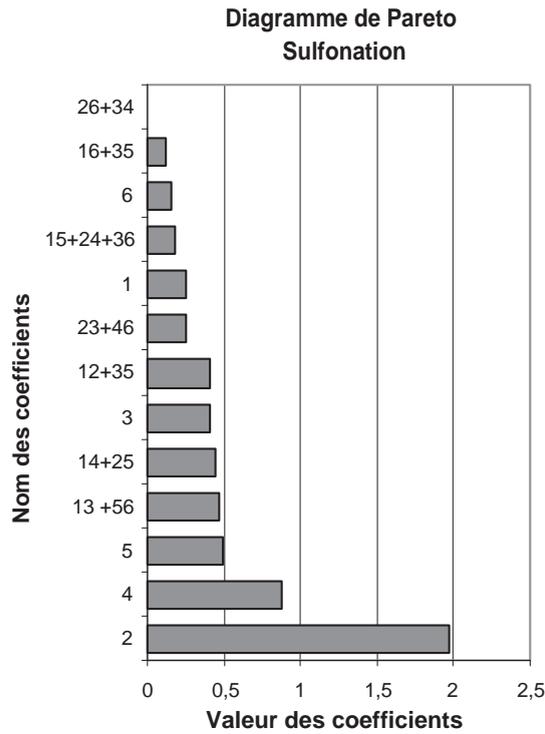


Figure 7.2 – Diagramme de Pareto des effets de l'exemple sur la sulfonation.

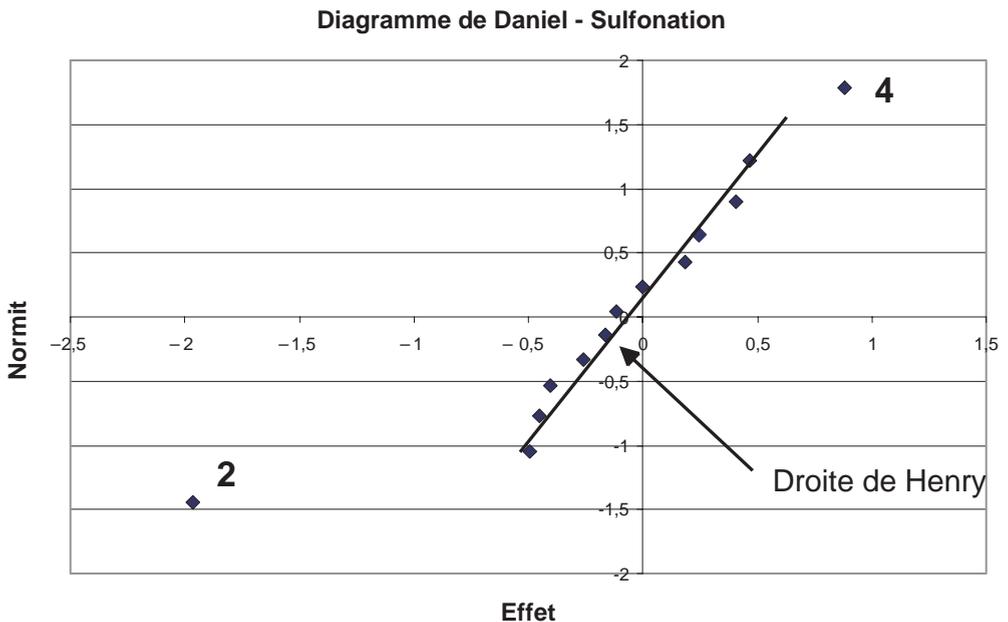


Figure 7.3 – Diagramme de Daniel permettant de distinguer les effets importants.

La décision est basée sur l'erreur de mesure et la connaissance des enjeux qui permet de définir le nombre d'écart-types à prendre en compte.

Quel que soit le raisonnement que vous utiliserez, il est toujours bon d'examiner le diagramme de Pareto et le diagramme de Daniel. Ils vous aident à prendre vos décisions.

L'examen des diagrammes de Pareto et de Daniel montre qu'il faut retenir un facteur influent : la concentration d'anhydride sulfurique dans l'acide sulfurique (2). Cet examen montre également que l'élimination de l'eau (4) est peu influente. La décision de le conserver ou de l'éliminer ne peut pas être basée sur une statistique. Ici, c'est un raisonnement chimique qui doit guider l'expérimentateur. Étant donné que l'élimination de l'eau modifie la concentration en acide sulfurique, il est chimiquement normal de considérer le facteur 4 comme légèrement influent et de le conserver. Les autres facteurs et les autres interactions sont très faibles et de peu d'intérêt pour résoudre le problème.

Dans un choix il y a toujours une part d'arbitraire, mais dans le cas des plans d'expériences, ce choix n'est jamais définitif et il peut être modifié pour donner lieu à une nouvelle analyse.

7.1.4 Conclusion de l'étude

L'importance relative des contrastes est indiquée par le diagramme à barres (Figure 7.1), le diagramme de Pareto (Figure 7.2) et le diagramme de Daniel (Figure 7.3).

Les contrastes ℓ_{12} , ℓ_{13} , ℓ_{14} , ℓ_{15} , ℓ_{16} , ℓ_{23} et ℓ_{26} sont faibles. D'après l'hypothèse 2, tous les coefficients aliasés dans ces contrastes sont nuls. Il n'y a donc pas d'interaction d'ordre 2 entre les facteurs.

Il ne reste que deux facteurs influents (Figure 7.4), la quantité d'anhydride sulfurique présente dans l'acide sulfurique (facteur 2) et l'élimination de l'eau en cours de réaction (facteur 4).

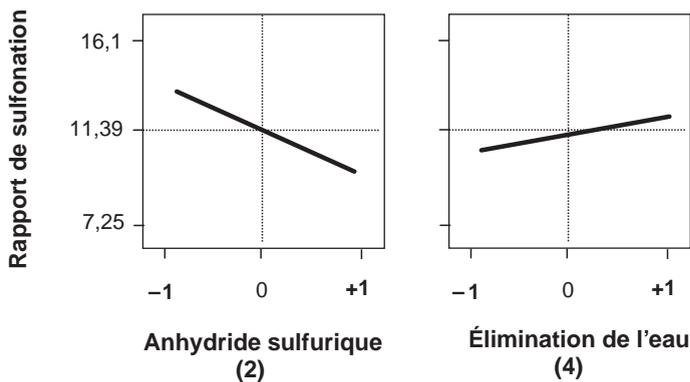


Figure 7.4 – Effets des deux facteurs influents.

La quantité d'anhydride sulfurique (SO_3) présente dans l'acide sulfurique ayant un effet négatif, la réponse sera élevée si l'on choisit le niveau bas, c'est-à-dire de ne pas mettre de SO_3 dans l'acide sulfurique.

L'élimination de l'eau en cours de réaction ayant un effet positif, la réponse sera élevée si l'on choisit le niveau haut, c'est-à-dire si on élimine l'eau en cours de réaction. Cet effet est peu élevé et pourrait être ignoré si l'on se référait simplement aux statistiques. Mais d'un point de vue chimique, l'élimination de l'eau augmente la concentration en acide sulfurique et déplace ainsi la réaction vers la formation d'acides sulfoniques. On ne peut pas négliger l'influence de ce facteur.

La sulfonation du benzène, en présence de toluène et de xylènes, est donc minimisée si l'on opère en milieu acide sulfurique concentré (mais sans anhydride sulfurique) en éliminant l'eau en cours de réaction.

7.2 Exemple 07 : le spectrofluorimètre

7.2.1 Préparation du plan d'expériences

■ Description de l'étude

Le benzopyrène est un produit catalogué comme cancérigène. Sa teneur dans les huiles lubrifiantes doit être inférieure à des quantités infinitésimales. Son analyse est donc délicate et nécessite des moyens importants. Les spécialistes chargés de mettre au point la méthode de dosage ont choisi la spectrofluorimétrie. La figure 7.5 présente le schéma de l'appareil : un faisceau lumineux issu d'une lampe à xénon tombe sur un premier monochromateur. La lumière est diffractée. Une lumière de longueur d'onde bien définie ressort du monochromateur en passant par une fente, la fente d'excitation. Ce rayon tombe sur l'échantillon qui en absorbe une partie et réémet dans toutes les directions une lumière de fluorescence. Cette lumière doit être analysée. Elle passe donc dans un deuxième monochromateur qui la résout en différentes longueurs d'onde. Puis elle ressort en passant à travers une fente, la fente d'émission, et elle est détectée par un photomultiplicateur. Enfin, le spectre correspondant est enregistré par un appareil muni d'un photomultiplicateur. La longueur d'onde d'un des pics du spectre est caractéristique du benzopyrène. Ce pic est situé à 481 nm et sa hauteur est proportionnelle à la concentration du benzopyrène. L'objectif de l'étude est d'obtenir un spectre de bonne qualité permettant de faire un dosage quantitatif du benzopyrène. Pour cela il faut que la hauteur du pic à 481 nm soit la plus élevée possible, il faut aussi que le petit pic situé à 489 nm soit bien détaché du pic à 481 nm (Figure 7.5). Enfin, le bruit de fond ne doit pas cacher le pic du benzopyrène à 481 nm quand la teneur est faible et que ce pic est petit.

Un tel appareil nécessite des réglages très fins pour que la concentration en benzopyrène soit déterminée avec fiabilité et précision.

■ Réponses

Les expérimentateurs ont défini trois réponses (Figure 7.6) :

- la *sensibilité* (A) évaluée par la hauteur du pic situé à 481 nm ;
- le *bruit de fond* (B) évalué par la distance entre l'enveloppe inférieure et l'enveloppe supérieure de la courbe d'enregistrement aux basses longueurs d'onde ;
- la *sélectivité* (C) évaluée par la largeur à mi-hauteur du pic à 481 nm.

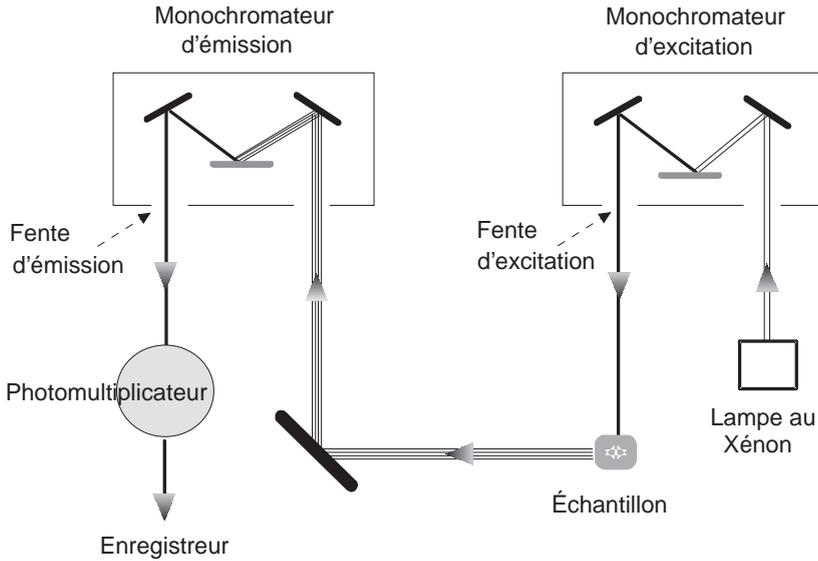


Figure 7.5 – Schéma d'un spectrofluorimètre.

Les réponses A et B sont exprimées en densité optique.

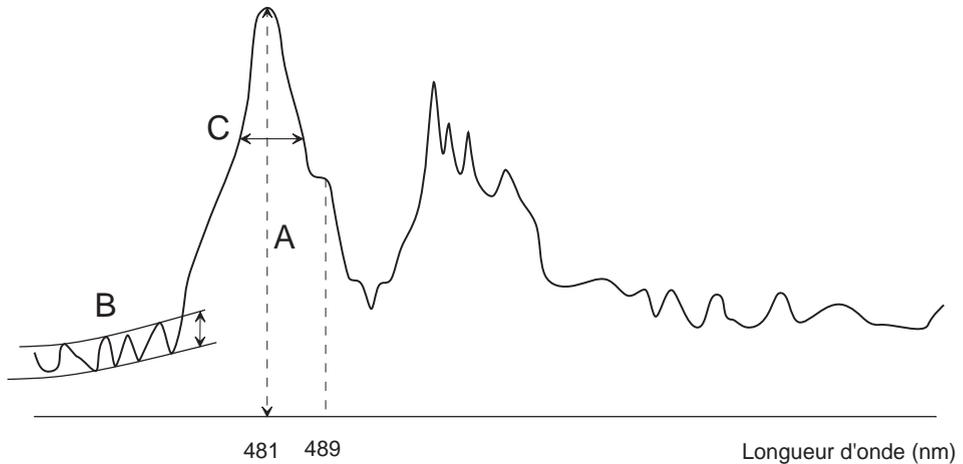


Figure 7.6 – Définition des réponses sur le spectre d'enregistrement : sensibilité (A), bruit de fond (B) et sélectivité (C).

■ Objectif de l'étude

Les cibles visées pour chacune des réponses sont les suivantes :

- Sensibilité : la hauteur du pic à 481 nm doit être la plus élevée possible.
- Sélectivité : la largeur à mi-hauteur du pic à 481 nm doit être la plus faible possible.
- Bruit de fond : la distance entre les enveloppes inférieure et supérieure doit être la plus faible possible.

■ Facteurs et domaine d'étude

D'après les experts en spectrofluorimétrie, sept facteurs peuvent modifier les réponses choisies :

- Facteur 1 : largeur de la fente d'excitation.
- Facteur 2 : largeur de la fente d'émission.
- Facteur 3 : température de l'échantillon.
- Facteur 4 : vitesse de balayage.
- Facteur 5 : gain de l'appareil.
- Facteur 6 : tension du photomultiplicateur.
- Facteur 7 : amortissement de la plume d'enregistrement.

Le domaine d'étude est un espace à 7 dimensions défini par les niveaux bas et haut de chacun des facteurs (Tableau 7.4).

Tableau 7.4 – Domaine d'étude.

Facteurs	Niveau -1	Niveau +1
Fente d'excitation (1)	2,5 nm	7,5 nm
Fente d'émission (2)	2,5 nm	7,5 nm
Température (3)	20 °C	40 °C
Vitesse balayage (4)	20	100
Gain (5)	1	10
Tension (6)	310 V	460 V
Amortissement (7)	2	4

■ Modélisation mathématique

On cherche les facteurs influents et leurs sens de variation ainsi que les interactions d'ordre 2.

Le modèle mathématique choisi *a priori* est un modèle polynomial classique du premier degré avec les interactions d'ordre 2. Ce modèle mathématique comporte 29 coefficients.

■ Choix du plan d'expériences

Avec 8 expériences, on peut déterminer 8 inconnues, c'est-à-dire 8 contrastes. Pour gagner notre pari, il faut que, pour chaque réponse, on puisse déterminer sans ambiguïtés les coefficients importants. Si ce n'est pas le cas, on risque d'avoir un plan complémentaire à effectuer.

On aliase le premier facteur supplémentaire, le facteur 4, sur l'interaction 123 du plan de base 2^3 . On peut écrire en notation de Box :

$$4 = 123$$

Puis, on aliase les facteurs suivants ainsi :

$$5 = 12$$

$$6 = 23$$

$$7 = 13$$

On en déduit les 4 générateurs d'alias indépendants :

$$I = 1234 = 125 = 236 = 137$$

L'expérimentateur ne prévoit pas de point de contrôle car il sait qu'il ne procédera à aucune modélisation pour faire des prévisions.

■ Construction du plan

Pour construire le plan 2^{7-4} , il suffit de prendre le plan de base 2^3 (Tableau 6.9 du chapitre 6) et d'attribuer les facteurs supplémentaires aux interactions.

Tableau 7.5 – Plan fractionnaire 2^{7-4} .

Essai n°	1	2	3	4 = 123	5 = 12	6 = 23	7 = 13
1	-	-	-	-	+	+	+
2	+	-	-	+	-	+	-
3	-	+	-	+	-	-	+
4	+	+	-	-	+	-	-
5	-	-	+	+	+	-	-
6	+	-	+	-	-	-	+
7	-	+	+	-	-	+	-
8	+	+	+	+	+	+	+

■ Confusions

Les quatre générateurs d'alias indépendants $I = 1234 = 125 = 236 = 137$ permettent de calculer les générateurs d'alias dépendants : on multiplie, au sens du calcul de Box, les générateurs d'alias indépendants deux par deux, trois par trois et quatre par quatre. On obtient :

$$1234 \times 125 = 345$$

$$1234 \times 236 = 146$$

$$1234 \times 137 = 247$$

$$125 \times 236 = 12356$$

$$125 \times 137 = 2357$$

$$236 \times 137 = 1267$$

$$1234 \times 125 \times 236 = 2456 \quad 1234 \times 125 \times 137 = 1457$$

$$1234 \times 236 \times 137 = 3467$$

$$125 \times 236 \times 137 = 567$$

$$1234 \times 125 \times 236 \times 137 = 1234567$$

Soit le générateur complet qui comprend 16 termes :

$$\mathbf{I} = 1234 = 125 = 236 = 137 = 345 = 146 = 247 = 12356 = 2357 \\ = 1267 = 2456 = 1457 = 3467 = 567 = 1234567$$

Pour obtenir les coefficients aliasés dans le contraste ℓ_1 , on multiplie les générateurs d'aliasés par **1**. Les générateurs perdent un chiffre s'ils possèdent un **1**, ou augmentent d'un chiffre s'ils ne possèdent pas de **1**. Comme nous avons supposé que les interactions d'ordre 3 étaient négligeables, nous pouvons éliminer du générateur complet les générateurs de plus de 4 chiffres. D'où le générateur complet simplifié :

$$\mathbf{I} = 234 = 125 = 236 = 137 = 345 = 146 = 247 = 567$$

Ce générateur complet simplifié permet de savoir comment sont aliasés les effets principaux et les interactions d'ordre 2 pour ce plan fractionnaire :

$$\ell_1 = a_1 + a_{25} + a_{37} + a_{46}$$

$$\ell_2 = a_2 + a_{15} + a_{36} + a_{47}$$

$$\ell_3 = a_3 + a_{17} + a_{26} + a_{45}$$

$$\ell_4 = a_4 + a_{16} + a_{27} + a_{35}$$

$$\ell_5 = a_5 + a_{12} + a_{34} + a_{67}$$

$$\ell_6 = a_6 + a_{14} + a_{23} + a_{57}$$

$$\ell_7 = a_7 + a_{13} + a_{24} + a_{56}$$

7.2.2 Réalisation des essais

Les essais sont réalisés en accord avec le plan d'expériences et les résultats sont notés dans le tableau 7.6.

Tableau 7.6 – Réponses du plan initial.

Essai n°	Sensibilité	Sélectivité	Bruit de fond
1	1,22	5,5	-1,47
2	0,9	9	-1,47
3	5,33	20	2,3
4	5,64	12	-0,69
5	3,89	7,5	0,69
6	3,88	8	0,4
7	2,82	13	0,26
8	2,33	23	-3,91

7.2.3 Interprétation des résultats du plan initial

■ Préparation des calculs

Un examen montre qu'il n'y a pas de corrélations marquées entre les trois réponses. On va donc les analyser les unes après les autres. On prépare les calculs en choisissant un modèle contenant les facteurs principaux et les interactions d'ordre 2. Examinons d'abord la sensibilité.

■ Réponse Sensibilité

On sait que les valeurs indiquées sont celles des contrastes (Tableau 7.7 et Figure 7.7). L'interprétation se fait donc en regardant la structure des aliases et en appliquant les règles d'interprétation.

Tableau 7.7 – Coefficients du modèle Sensibilité (unités codées).

Effet	Valeur
ℓ_0 (constante)	3,25
ℓ_1	-0,06
ℓ_2	0,78
ℓ_3	-0,02
ℓ_4	-0,14
ℓ_5	0,018
ℓ_6	-1,43
ℓ_7	-0,06

Deux contrastes sont forts : ℓ_2 et ℓ_6 .

Il faut se méfier de l'interaction 26 (hypothèse 3). L'interaction 26 est aliasée avec l'effet du facteur 3, l'interaction 17 et l'interaction 45 dans le contraste ℓ_3 . Le contraste ℓ_3 est pratiquement nul. D'après l'hypothèse 2, tous les coefficients de ce contraste sont nuls. L'interaction 26 peut donc être négligée.

S'il n'y avait que la réponse Sensibilité, l'étude serait terminée et l'on saurait qu'il n'y a que deux facteurs influents : la fente d'émission (facteur 2) et la tension du photomultiplicateur (facteur 6). On peut vérifier cette conclusion en choisissant un modèle avec ces deux facteurs et leur interaction. Ce modèle possède un coefficient de détermination de 0,99. Le diagramme des effets indique que les niveaux qui favorisent une forte sensibilité sont le niveau haut pour le facteur 2 et le niveau bas pour le facteur 6 (Figure 7.8).

Les alias de la Sensibilité

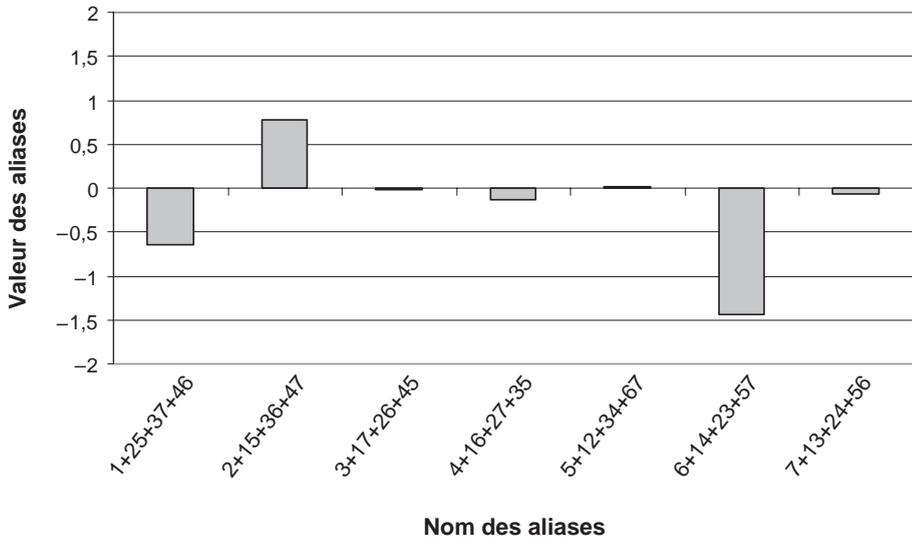


Figure 7.7 – Diagramme à barres des alias (sensibilité).

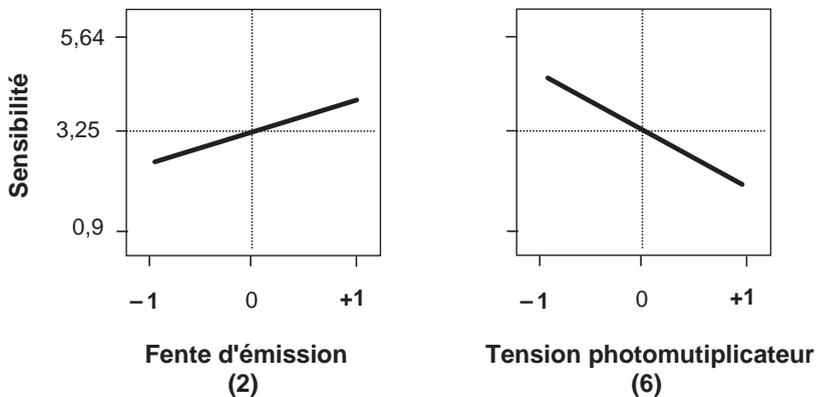


Figure 7.8 – Diagramme des effets des facteurs influents sur la sensibilité.

■ Réponse Bruit de fond

On mène les calculs de la même manière que pour la sensibilité, c'est-à-dire que l'on considère un modèle avec toutes les interactions d'ordre 2. La structure des alias est la même que pour la sensibilité. L'analyse des résultats se conduit de la même manière.

- Trois contrastes sont forts : ℓ_1 , ℓ_5 et ℓ_6 (Tableau 7.8 et Figure 7.9).
- Il faut donc se méfier des interactions 15, 16 et 56 (hypothèse 3).
- L'interaction 15 est aliassée avec l'effet du facteur 2, l'interaction 36 et l'interaction 47 dans le contraste ℓ_2 . Le contraste ℓ_2 est pratiquement nul. D'après

Tableau 7.8 – Coefficients du modèle Bruit de fond (unités codées).

Effet	Valeur
ℓ_0 (constante)	-0,486
ℓ_1	-0,931
ℓ_2	-0,024
ℓ_3	-0,154
ℓ_4	-0,111
ℓ_5	-0,858
ℓ_6	-1,161
ℓ_7	-0,183

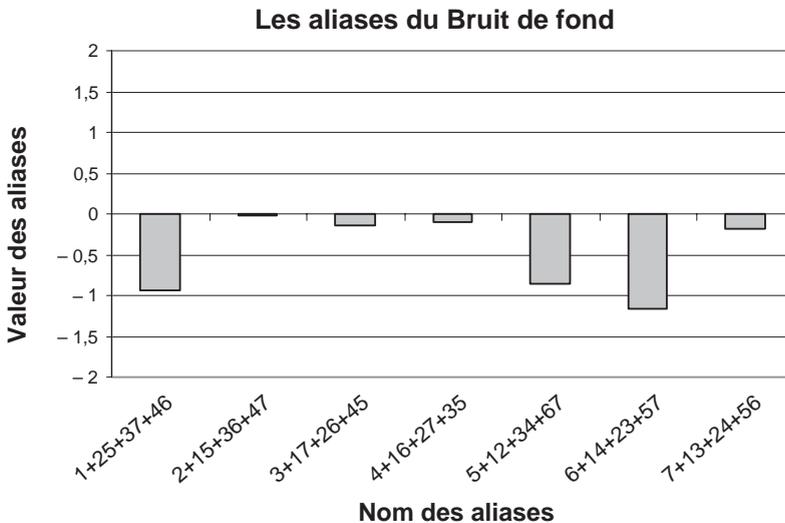


Figure 7.9 – Diagramme à barres des aliases (bruit de fond).

l'hypothèse 2, tous les coefficients de ce contraste sont nuls. L'interaction 15 peut donc être négligée.

- L'interaction 16 est aliassée avec l'effet du facteur 4, l'interaction 27 et l'interaction 35 dans le contraste ℓ_4 . Le contraste ℓ_4 est pratiquement nul. D'après l'hypothèse 2, tous les coefficients de ce contraste sont nuls. L'interaction 16 peut donc être négligée.
- L'interaction 56 est aliassée avec l'effet du facteur 7, l'interaction 13 et l'interaction 24 dans le contraste ℓ_7 . Le contraste ℓ_7 est pratiquement nul. D'après l'hypothèse 2, tous les coefficients de ce contraste sont nuls. L'interaction 56 peut donc être négligée.

Le bruit de fond est influencé par la fente d'excitation (facteur 1), par le gain de l'appareil (facteur 5) et par la tension du photomultiplicateur (facteur 6). Il n'y a pas d'interactions entre ces facteurs.

On peut vérifier cette conclusion en choisissant un modèle avec ces trois facteurs et leurs interactions. Le modèle a un coefficient de détermination de 0,99. Les niveaux bas des trois facteurs Fente d'excitation (1), Gain de l'appareil (5) et Tension du photomultiplicateur (6), favorisent un faible bruit de fond (Figure 7.10).

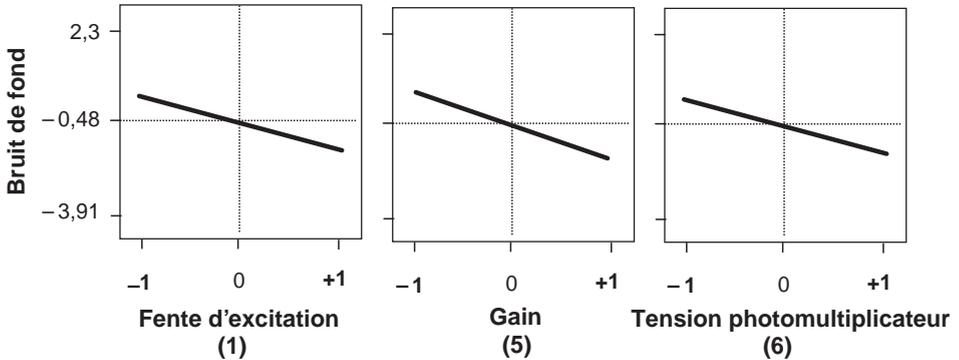


Figure 7.10 – Diagramme des effets des facteurs influents sur le bruit de fond.

S'il n'y avait que les deux réponses Sensibilité et Bruit de fond, l'étude serait terminée et l'on saurait qu'il y a :

- deux facteurs influents sur la sensibilité : la fente d'émission (facteur 2) et la tension du photomultiplicateur (facteur 6). Il n'y a pas d'interaction entre ces deux facteurs ;
- trois facteurs influents sur le bruit de fond : la fente d'excitation (facteur 1), le gain de l'appareil (facteur 5) et la tension du photomultiplicateur (facteur 6). Il n'y a pas d'interaction entre ces trois facteurs.

■ Réponse Sélectivité

Les calculs sont menés de la même manière que pour la sensibilité et le bruit de fond, c'est-à-dire que l'on prend un modèle avec toutes les interactions d'ordre 2. La structure des alias est donc la même que pour la sensibilité et le bruit de fond. L'analyse des résultats se conduit de la même manière.

- Trois contrastes sont forts : ℓ_2 , ℓ_4 et ℓ_7 (Tableau 7.9 et Figure 7.11).
- Il faut donc se méfier des interactions 24, 27 et 47 (hypothèse 3).
- L'interaction 24 est aliassée avec l'effet du facteur 7, l'interaction 13 et l'interaction 56 dans le contraste ℓ_7 . Le contraste ℓ_7 n'est pas faible et, par conséquent, on ne peut pas appliquer l'hypothèse 2. On ne peut pas savoir si le contraste ℓ_7 est fort à cause de l'effet du facteur 7 ou de l'interaction 24. Il y a ambiguïté.
- L'interaction 27 est aliassée avec l'effet du facteur 4, l'interaction 16 et l'interaction 35 dans le contraste ℓ_4 . Le contraste ℓ_4 ne peut pas être négligé et donc

Tableau 7.9 – Coefficients du modèle Sélectivité (unités codées).

Effet	Valeur
ℓ_0 (constante)	12,25
ℓ_1	0,75
ℓ_2	4,75
ℓ_3	0,625
ℓ_4	2,625
ℓ_5	-0,25
ℓ_6	0,375
ℓ_7	1,875

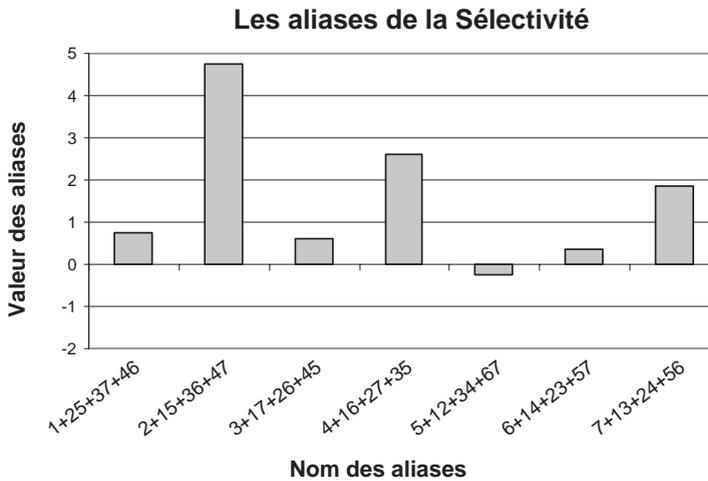


Figure 7.11 – Diagramme à barres des aliases (sélectivité).

l'hypothèse 2 ne s'applique pas. On ne peut pas savoir si le contraste ℓ_4 est fort à cause de l'effet du facteur 4 ou de l'interaction 27. Il y a ambiguïté.

- L'interaction 47 est aliéée avec l'effet du facteur 2, l'interaction 15 et l'interaction 36 dans le contraste ℓ_2 . Le contraste ℓ_2 ne peut pas être négligé et donc l'hypothèse 2 ne s'applique pas. On ne peut pas savoir si le contraste ℓ_2 est fort à cause de l'effet du facteur 2 ou de l'interaction 47. Il y a ambiguïté.

La réponse Sélectivité nous oblige à envisager un plan complémentaire pour désaliéer :

- l'effet du facteur 2 (fente d'émission) de l'interaction 47 ;
- l'effet du facteur 4 (vitesse balayage) de l'interaction 27 ;
- l'effet du facteur 7 (amortissement) de l'interaction 24.

■ Bilan provisoire des résultats acquis

Avant d'attaquer le plan complémentaire, on peut faire le point sur les renseignements déjà acquis. Dans le tableau 7.10, on indique les niveaux qui sont favorables aux objectifs qui ont été définis.

Tableau 7.10 – Niveaux favorables des facteurs influents sur les réponses.

Facteurs	Sensibilité	Bruit de fond	Sélectivité
Fente d'excitation (1)		+	
Fente d'émission (2)	+		?
Température (3)			
Vitesse balayage (4)			?
Gain (5)		+	
Tension (6)	-	+	
Amortissement (7)			?

On constate que :

- Le facteur 1 n'agit que sur le bruit de fond.
- Le facteur 2 agit sur la sensibilité et peut-être sur la sélectivité.
- Le facteur 3 n'a aucune influence sur les réponses.
- Si le facteur 4 agit, c'est uniquement sur la sélectivité.
- Le facteur 5 n'agit que sur le bruit de fond.
- Le facteur 6 agit sur la sensibilité et sur le bruit de fond. Les niveaux de réglage sont incompatibles. Cela n'est pas pour surprendre car si les réglages augmentent (ou diminuent) la densité optique, ils augmenteront (ou diminueront) aussi bien le pic à 481 nm que le bruit de fond.
- Si le facteur 7 agit, c'est uniquement sur la sélectivité.

7.2.4 Éléments pour construire le plan complémentaire

Le plan complémentaire doit permettre d'obtenir l'effet du facteur 2 sans les interactions 15, 36 et 47, c'est-à-dire qu'il faut trouver le plan ayant comme contraste :

$$\ell'_2 = a_2 - a_{15} - a_{36} - a_{47}$$

Ce contraste correspond aux confusions :

$$2 = -15 = -36 = -47$$

En multipliant par 2, on obtient trois générateurs du plan cherché :

$$I = -125 = -236 = -247$$

Comme il faut quatre générateurs, on reprend :

$$I = 1234$$

Pour associer le générateur -247 avec une interaction du plan de base, on peut remplacer 4 par 123 :

$$-247 = -21237 = -137$$

D'où les quatre générateurs indépendants du plan complémentaire :

$$I = -125 = -236 = -137 = 1234$$

On en déduit les colonnes de signes du plan complémentaire. On applique le calcul de Box en multipliant les générateurs indépendants par 4 , 5 , 6 et 7 :

$$4 = 123$$

$$5 = -12$$

$$6 = -23$$

$$7 = -13$$

Il faut donc changer les signes des colonnes des facteurs 5, 6 et 7.

Vérifions que ce choix désaliasie bien l'effet du facteur 2 de l'interaction 47. Le plan initial fournit un contraste ℓ_2 dont la structure est :

$$\ell_2 = a_2 + a_{15} + a_{36} + a_{47}$$

Le plan complémentaire fournit un contraste ℓ'_2 dont la structure est :

$$\ell'_2 = a_2 - a_{15} - a_{36} - a_{47}$$

L'analyse des deux plans donnera la somme et la différence de ces deux contrastes : on obtiendra d'une part l'effet du facteur 2 seul, d'autre part la somme des trois interactions :

$$\frac{\ell_2 + \ell'_2}{2} = a_2 \qquad \frac{\ell_2 - \ell'_2}{2} = a_{15} + a_{36} + a_{47}$$

L'effet du facteur 2 est bien désaliasé de l'interaction 47.

Ce qui est vrai pour le facteur 2, est vrai pour les autres facteurs : d'une manière générale, les effets des facteurs seront désaliasés des interactions d'ordre 2 et celles-ci resteront aliasées entre elles.

■ Construction du plan complémentaire

Pour construire le plan complémentaire, on change simplement les signes des interactions 12, 13 et 23.

Tableau 7.11 – Plan fractionnaire complémentaire 2^{7-4} .

Essai n°	1	2	3	4 = 123	5 = -12	6 = -23	7 = -13
9	-	-	-	-	-	-	-
10	+	-	-	+	+	-	+
11	-	+	-	+	+	+	-
12	+	+	-	-	-	+	+
13	-	-	+	+	-	+	+
14	+	-	+	-	+	+	-
15	-	+	+	-	+	-	+
16	+	+	+	+	-	-	-

On exécute ces 8 nouveaux essais et on mesure les 3 réponses.

Tableau 7.12 – Réponses du plan complémentaire.

Essai n°	Sensibilité	Sélectivité	Bruit de fond
9	4,14	3,5	3,56
10	3,18	12	-2,41
11	2,82	14	1,39
12	2,74	14	-1,9
13	2,44	8	-0,62
14	0,98	6	-3,22
15	5,66	14	0,02
16	5,63	14	-1,61

■ Confusions (plan initial et plan complémentaire)

On possède maintenant un plan 2^{7-3} de 16 essais dont le générateur indépendant est :

$$I = 1234 = 3467 = 1356 = 1457 = 1267 = 1356 = 2357$$

On voit que les facteurs principaux sont aliasés avec des interactions d'ordre 3 ou plus élevé. Comme d'après l'hypothèse 1, ces interactions sont négligeables, les contrastes des effets principaux sont égaux aux effets.

Les interactions d'ordre 2 sont aliasées entre elles trois par trois :

$$\ell_{12} = a_{12} + a_{34} + a_{67}$$

$$\ell_{13} = a_{13} + a_{24} + a_{56}$$

$$\ell_{14} = a_{14} + a_{23} + a_{57}$$

$$\ell_{15} = a_{15} + a_{36} + a_{47}$$

$$\ell_{16} = a_{16} + a_{35} + a_{27}$$

$$\ell_{17} = a_{17} + a_{26} + a_{45}$$

$$\ell_{25} = a_{25} + a_{37} + a_{46}$$

En particulier l'effet du facteur 2 est désaliasé de l'interaction 47, l'effet du facteur 4 est désaliasé de l'interaction 27 et l'effet du facteur 7 est désaliasé de l'interaction 24.

On va d'abord analyser la sélectivité pour savoir si les deux plans réunis sont suffisants pour expliquer cette réponse. Si oui, on pourra s'arrêter, sinon on continuera l'expérimentation en ajoutant les essais qui permettront de lever les nouvelles ambiguïtés.

On vérifie ensuite que les analyses faites sur les réponses Sensibilité et Bruit de fond sont confirmées sur les 16 essais.

■ Les trois réponses analysées sur seize essais

On reprend l'ensemble des 16 essais et on refait les calculs. On obtient les coefficients, et non les contrastes, pour les effets principaux. Les interactions sont regroupées dans les contrastes (Tableau 7.13).

Tableau 7.13 – Coefficients (sensibilité, bruit de fond, sélectivité).

Effet	Sensibilité	Bruit de fond	Sélectivité
Constante	3,35	-0,54	11,47
Fente d'excitation (1)	-0,19	-1,31	0,78
Fente d'émission (2)	0,77	0,02	4,03
Température (3)	0,10	-0,46	0,22
Vitesse balayage (4)	-0,035	-0,16	1,97
Gain (5)	-0,13	-0,66	0,28
Tension (6)	-1,32	-0,82	0,09
Amortissement (7)	-0,002	-0,41	1,59
$\ell_{12} = 12 + 34 + 67$	0,15	-0,20	-0,53
$\ell_{13} = 13 + 24 + 56$	-0,058	0,22	0,28
$\ell_{14} = 14 + 23 + 57$	-0,115	-0,34	0,28
$\ell_{15} = 15 + 36 + 47$	0,007	-0,05	0,72
$\ell_{16} = 16 + 27 + 35$	-0,10	0,05	0,66
$\ell_{17} = 17 + 26 + 45$	-0,12	0,30	0,41
$\ell_{25} = 25 + 37 + 46$	0,126	0,38	-0,03

Les expérimentateurs ont considéré que toutes les interactions étaient insuffisamment importantes pour modifier de manière intéressante les réglages de l'appareil (Figure 7.12). Dans ce cas, seuls les effets principaux sont à prendre en compte. Il n'y a donc que trois facteurs influents sur la sélectivité et il n'y a pas d'interactions importantes entre ces facteurs.

La sélectivité, qui est mesurée par la largeur du pic à mi-hauteur, sera bonne si cette largeur est petite. Il faut donc régler les facteurs 2, 4 et 7 aux niveaux bas (Figure 7.13). Pour la sensibilité, on retrouve bien les deux facteurs importants : la fente d'émission (facteur 2) et la tension du photomultiplicateur (facteur 6). Il n'y a pas d'interactions

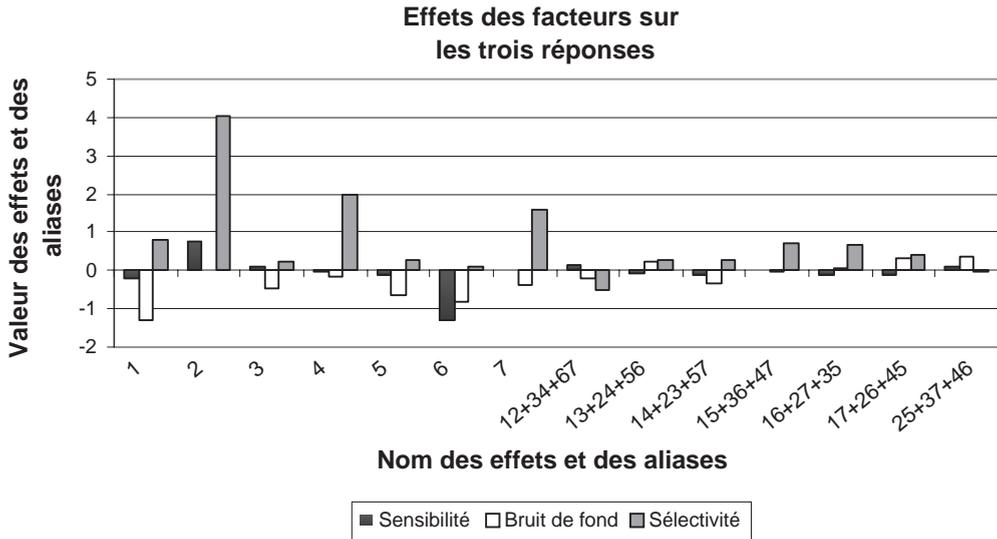


Figure 7.12 – Diagramme à barres des effets et des aliasses des trois réponses.

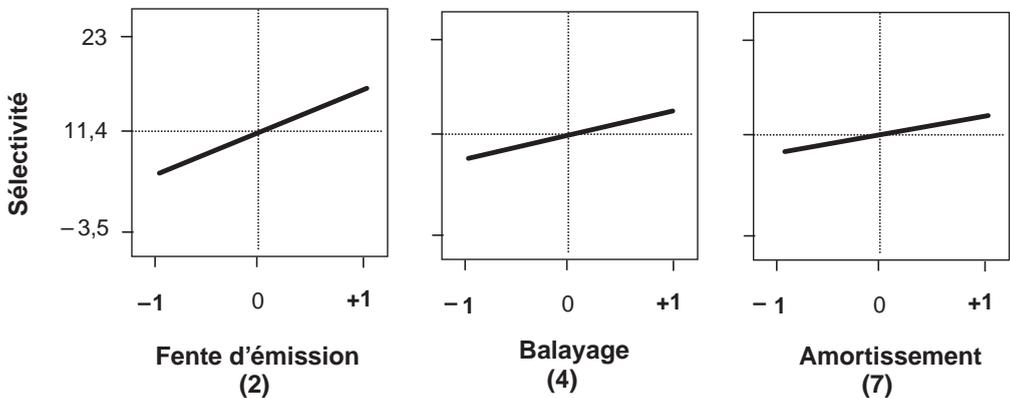


Figure 7.13 – Diagramme des effets des facteurs influents sur la sélectivité.

importantes entre ces facteurs. Pour obtenir une bonne sensibilité, il faut régler le facteur 2 au niveau haut et le facteur 6 au niveau bas.

Pour le bruit de fond, on retrouve bien que les trois facteurs importants sont la fente d'excitation (facteur 1), le gain du photomultiplicateur (facteur 5) et la tension du photomultiplicateur (facteur 6). Il n'y a pas d'interactions importantes entre ces facteurs. Pour obtenir un faible bruit de fond, il faut régler les facteurs 1, 5 et 6 au niveau haut.

7.2.5 Conclusion de l'étude

Nous devons rassembler l'ensemble des conclusions obtenues sur chaque réponse pour faire le bilan de l'interprétation. Les résultats que nous avons consignés dans

le tableau 7.10 sont confirmés pour la sensibilité et le bruit de fond. Nous pouvons maintenant le compléter pour la sélectivité (Tableau 7.14).

Tableau 7.14 – Niveaux favorables des facteurs influents sur les réponses.

Facteurs	Sensibilité	Bruit de fond	Sélectivité
Fente d'excitation (1)		+	
Fente d'émission (2)	+		-
Température (3)			
Vitesse balayage (4)			-
Gain (5)		+	
Tension (6)	-	+	
Amortissement (7)			-

Sur les 7 facteurs, 5 sont faciles à régler :

- La fente d'excitation (facteur 1) doit être réglée à 7,5 nm pour diminuer le bruit de fond.
- La température de l'échantillon (facteur 3) n'a aucune influence sur les réponses. Elle peut donc varier entre 20 et 40 °C.
- La vitesse de balayage (facteur 4) doit être réglée à 20 pour réduire la largeur du pic à mi-hauteur (sélectivité).
- Le gain (facteur 5) doit être réglé sur 10 pour abaisser le bruit de fond.
- L'amortissement (facteur 7) doit être réglé sur 2 pour favoriser la sélectivité.

Deux facteurs sont délicats à régler :

- Si la tension du photomultiplicateur (facteur 6) est réglée sur 310 V, la sensibilité sera élevée mais le bruit de fond sera important. Inversement si la tension du photomultiplicateur est de 460 V, il y aura peu de bruit de fond mais la sensibilité sera diminuée. Il faudra trouver le meilleur compromis entre ces deux réponses.
- Une large fente d'émission (facteur 2) augmente la sensibilité mais dégrade la sélectivité. Une fente d'émission étroite améliore la sélectivité, mais dégrade la sensibilité. Il faudra régler la largeur de cette fente pour trouver le meilleur compromis entre ces deux réponses.

Un spectre a été réalisé avec ces réglages (Figure 7.14). Le pic à 481 nm est élevé et se détache bien de celui situé à 489 nm, le bruit de fond a pratiquement disparu.

Dans la pratique, on pourra régler les cinq premiers facteurs aux niveaux conseillés et réaliser un plan d'optimisation avec les deux autres facteurs, tension du photomultiplicateur et fente d'émission, pour trouver le meilleur compromis en fonction de la concentration en benzopyrène et de la présence d'impuretés dans les solutions à analyser.

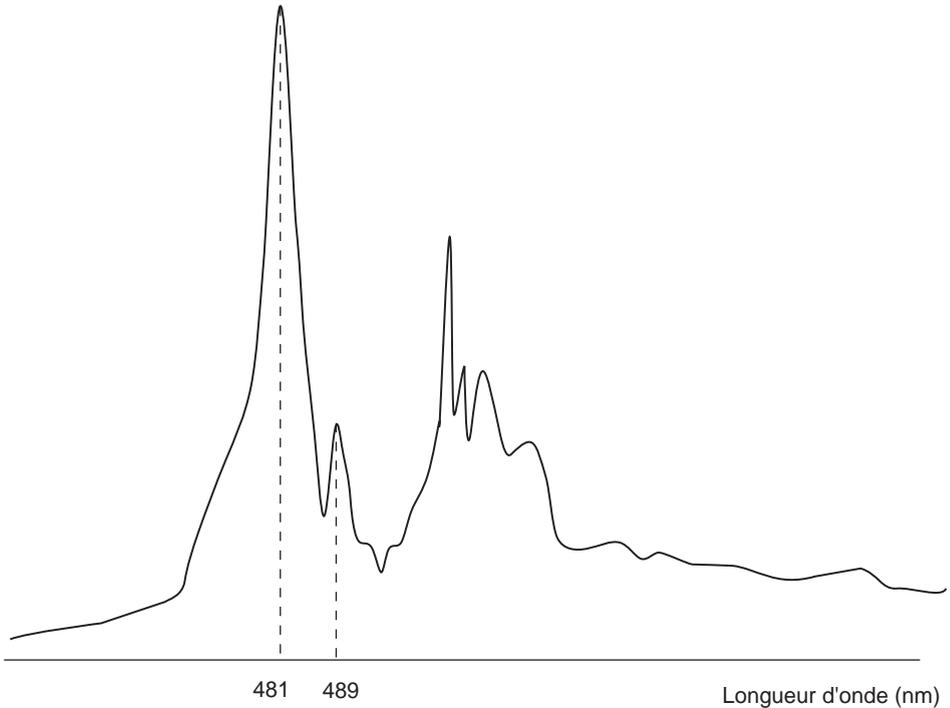


Figure 7.14 – Spectre enregistré avec les conditions optimales définies à l'aide du plan d'expériences.

7.3 Exemple 08 : les pommes de terre chips

7.3.1 Préparation du plan d'expériences

■ Description de l'étude

Les chips sont de fines tranches de pommes de terre que l'on laisse frire dans un bain d'huile. Ces chips sont agréables à manger à condition de ne pas trop sentir l'huile. À la fin de la cuisson la plus grosse partie de l'huile est extraite sous une pression de 40 MPa pendant 1 heure à 60 °C. Puis les chips sont soumises à un traitement de finition qui élimine encore de l'huile. Il s'agit d'un traitement sous une pression de 10 MPa pendant 30 min sous un débit de 700 kg/h de gaz carbonique supercritique. Enfin les chips sont salées et emballées. Il y a plusieurs types de chips, des petites ou des grandes et des plates ou des striées.

■ Réponses

Les responsables de la fabrication ont défini deux réponses :

- la quantité d'huile retirée qui est évaluée en poids,
- le goût des chips qui est évalué par un jury constitué d'experts qui mettent la note 0 si la qualité est mauvaise et 10 si elle est bonne.

■ Objectif de l'étude

L'objectif de l'étude est de sélectionner les facteurs les plus influents sur les deux réponses choisies. Il s'agit d'une étude préliminaire devant servir à préparer un plan de modélisation. Il s'agit donc d'une étude de criblage faisant intervenir un grand nombre de facteurs.

■ Facteurs et domaine d'étude

Les facteurs retenus couvrent l'ensemble du processus de déshuilage depuis l'extraction jusqu'au salage. Neuf facteurs pouvant être influents ont été recensés :

- Facteur 1 : pression d'extraction de l'huile.
- Facteur 2 : température d'extraction de l'huile.
- Facteur 3 : durée d'extraction de l'huile.
- Facteur 4 : pression de séparation de l'huile.
- Facteur 5 : température de séparation de l'huile.
- Facteur 6 : débit du gaz carbonique supercritique.
- Facteur 7 : quantité de sel ajoutée aux chips.
- Facteur 8 : taille des chips.
- Facteur 9 : type de chips.

Le domaine d'étude est défini par les niveaux bas et haut de chacun des facteurs (Tableau 7.15).

Tableau 7.15 – Domaine d'étude.

Facteurs	Unités	Niveau -1	Niveau +1
Pression d'extraction (1)	MPa	30	50
Température d'extraction (2)	°C	50	70
Durée d'extraction (3)	min	45	75
Pression de séparation (4)	MPa	5	15
Température de séparation (5)	°C	25	40
Débit CO ₂ (6)	kg/h	600	800
Sel (7)	g	0	1
Taille (8)		petite	grande
Type (9)		plat	strié

■ Modélisation mathématique

On cherche les facteurs influents et leurs sens de variation. Le modèle mathématique postulé est un simple modèle polynomial ne faisant intervenir que les termes du premier degré :

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5 + a_6x_6 + a_7x_7 + a_8x_8 + a_9x_9 \quad \{7.1\}$$

On adopte le même modèle pour les deux réponses.

■ Choix du plan d'expériences

Il y a 10 inconnues à déterminer : la constante et les 9 effets principaux. Il faut donc au moins 10 essais. Le plan classique le plus proche est un plan de Plackett et Burman de 12 essais.

Ces plans peuvent se construire à la main en suivant les règles énoncées par Plackett et Burman, mais il est plus commode de faire appel à un logiciel de plans d'expériences. Ces plans comportent un nombre d'essais multiple de 4. Les plans factoriels 2^k comportent également un nombre d'essais multiple de 4. Cela provient du fait que tous ces plans à deux niveaux sont construits sur des matrices d'Hadamard qui comportent un nombre de lignes (ou de colonnes) multiple de 4. Ces matrices d'Hadamard possèdent l'importante propriété d'orthogonalité, c'est-à-dire que le produit scalaire de deux colonnes quelconques est égal à 0.

L'habitude a été prise d'appeler *plan factoriel* (complet ou fractionnaire) tous les plans ayant un nombre d'essais égal à 2^k et d'appeler tous les autres *plans de Plackett et Burman*. Ainsi un plan de 12 essais est un plan de Plackett et Burman mais un plan de 16 essais est un plan factoriel. Cette distinction se justifie car les propriétés de ces plans sont un peu différentes. Par exemple, les alias des plans factoriels peuvent se calculer avec le calcul de Box mais les alias des plans de Plackett et Burman ne peuvent pas se calculer avec le calcul de Box.

Tableau 7.16 – Plan d'expériences de Plackett et Burman de 12 essais.

Essai n°	PE	Te	DE	PS	TS	CO ₂	Sel	Taille	Type
1	1	-1	1	-1	-1	-1	1	1	-1
2	1	1	-1	1	-1	-1	-1	1	-1
3	-1	1	1	-1	1	-1	-1	-1	-1
4	1	-1	1	1	-1	1	-1	-1	1
5	1	1	-1	1	1	-1	1	-1	1
6	1	1	1	-1	1	1	-1	1	1
7	-1	1	1	1	-1	1	1	-1	-1
8	-1	-1	1	1	1	-1	1	1	1
9	-1	-1	-1	1	1	1	-1	1	-1
10	1	-1	-1	-1	1	1	1	-1	-1
11	-1	1	-1	-1	-1	1	1	1	1
12	-1	-1	-1	-1	-1	-1	-1	-1	1

Tableau 7.16 – Plan d'expériences de Plackett et Burman de 12 essais. (Suite)

Essai n°	PE	Te	DE	PS	TS	CO ₂	Sel	Taille	Type
-	30	50	45	5	25	600	0	petite	plat
+	50	70	75	15	40	800	1	grande	strié

7.3.2 Réalisation des essais

Les résultats de l'expérimentation figurent dans le tableau 7.17.

Tableau 7.17 – Réponses du plan.

Essai n°	Huile	Goût
1	50,4	8
2	53,9	1
3	45,8	5
4	55,4	1
5	53,3	4
6	50,1	2
7	50,4	7
8	49,8	7
9	49,2	3
10	50,6	7
11	44,9	9
12	45,7	5

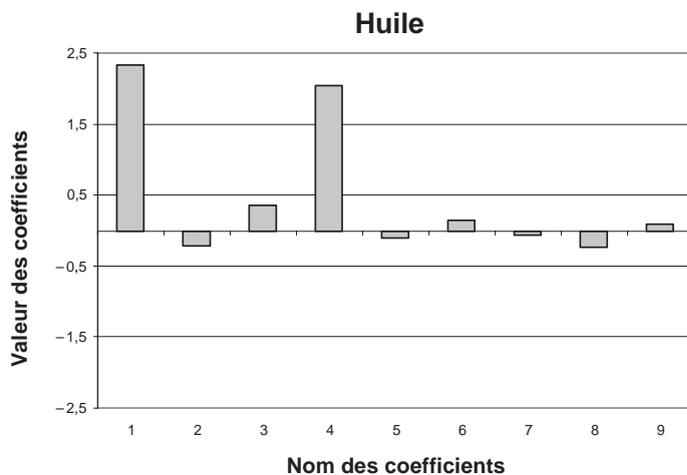
7.3.3 Interprétation des résultats du plan et conclusion

■ Réponse Huile

À partir des résultats expérimentaux obtenus sur l'huile, on calcule la valeur des coefficients de chaque facteur en utilisant le modèle simplifié {7.1}. Ces valeurs figurent dans le tableau 7.18 et sont illustrées par la figure 7.15. Deux facteurs sont nettement influents : la pression d'extraction (1) et la pression de séparation (4). Les autres facteurs ne jouent aucun rôle important sur l'élimination de l'huile.

Tableau 7.18 – Coefficients du modèle Huile (unités codées).

Effet	Valeur	p-value
Constante	49,96	< 0,0001
Pression d'extraction (1)	2,32	0,0025
Température d'extraction (2)	-0,22	0,1943
Durée d'extraction (3)	0,36	0,0921
Pression de séparation (4)	2,04	0,0033
Température de séparation (5)	-0,16	0,3085
Débit CO ₂ (6)	0,14	0,3495
Sel (7)	-0,06	0,6674
Taille (8)	-0,24	0,1748
Type (9)	-0,09	0,5153

**Figure 7.15** – Coefficients du modèle de la réponse Huile.

Les effets étant positifs, on extrait plus d'huile à forte pression qu'à faible pression.

■ Réponse Goût

La réponse Goût est traitée de la même manière que la réponse Huile. Le tableau 7.19 et la figure 7.16 montrent qu'il y a un seul facteur très influent, le sel (7), et deux facteurs influents, la pression d'extraction (1) et la pression de séparation (4).

Tableau 7.19 – Coefficients du modèle Goût (unités codées).

Effet	Valeur	p-value
Constante	4,9	0,0003
Pression d'extraction (1)	-1,08	0,0059
Température d'extraction (2)	-0,25	0,0955
Durée d'extraction (3)	0,08	0,4226
Pression de séparation (4)	-1,08	0,0059
Température de séparation (5)	-0,25	0,0955
Débit CO ₂ (6)	-0,08	0,4226
Sel (7)	2,08	0,0016
Taille (8)	0,08	0,4226
Type (9)	-0,25	0,0955

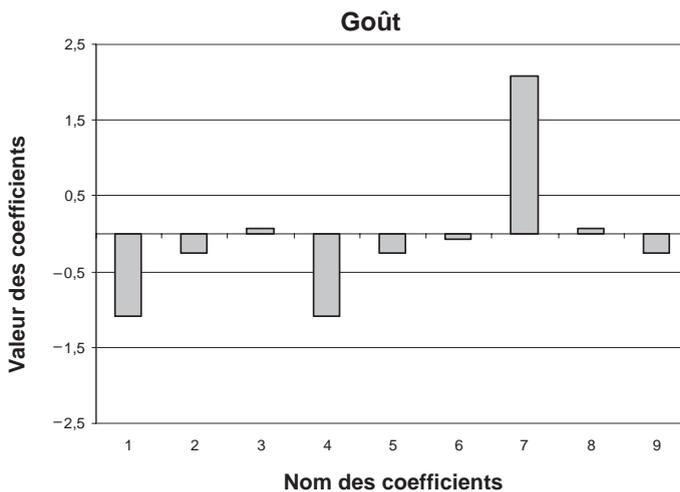


Figure 7.16 – Coefficients du modèle de la réponse Goût.

On améliore le goût si l'on met du sel.

Les chips sont donc meilleures au niveau bas des pressions d'extraction et de séparation. Or c'est pour ce niveau que l'on retire le moins d'huile. Les chips ont donc moins de goût quand on retire trop d'huile. On voit bien la difficulté de cette fabrication : s'il y a trop d'huile les chips sont grasses et s'il n'y a pas assez d'huile elles manquent de moelleux. La quantité d'huile qui reste sur les chips doit donc être bien équilibrée. On comprend pourquoi un plan d'optimisation est nécessaire : il faut trouver le bon équilibre entre le sel et l'huile restante dans les chips. Un plan faisant intervenir les trois facteurs, le sel, la pression d'extraction et la pression de séparation, devrait apporter la solution cherchée. À travers cet exemple, on voit tout l'intérêt d'un plan de criblage qui élimine tous les facteurs ne jouant qu'un rôle mineur sur les réponses étudiées.

8 • ORDRE DES ESSAIS

Est-il indifférent de réaliser les essais dans un ordre ou dans un autre ? Pour répondre à cette question, il faut tenir compte de plusieurs éléments, les uns pratiques, les autres théoriques. Le choix de l'ordre des essais s'effectuera en tenant compte de ces deux aspects.

Des impératifs expérimentaux peuvent imposer un ordre. Par exemple, une pièce en verre doit prendre deux positions dans un montage d'étude. Pour éviter de la casser en la déplaçant plusieurs fois, il peut être avantageux de réaliser la séquence des essais des niveaux bas puis, ensuite, la séquence des essais des niveaux hauts. On ne la déplacera ainsi qu'une fois. La mise en équilibre thermique d'un four peut être très longue à établir et l'expérimentateur ne tient pas à la refaire plusieurs fois. On pourrait trouver d'autres exemples où les conditions mêmes des expériences imposent un ordre de réalisation des essais, qu'il s'agisse de contraintes matérielles, temporelles ou autres. L'expérimentateur peut soupçonner une évolution dans le temps ou dans l'espace du phénomène qu'il étudie et donc des variations systématiques de la réponse ou des réponses. Il devra en tenir compte. Si l'évolution est régulière, par exemple l'usure d'une pièce mécanique, ou le vieillissement d'un produit, il y a dérive et le plan d'expériences devra être ordonné pour la mesurer ou l'écarter des résultats importants. On peut utiliser des plans anti-dérive et/ou ajouter des points de contrôle pour suivre l'évolution de cette dérive.

Il se peut que le lieu ou le moment où l'on effectue les essais affecte les résultats de l'expérimentation. Par exemple, les mesures effectuées le matin peuvent être systématiquement plus fortes que celles de l'après-midi. L'expérimentateur devra mener ses expériences en conséquence. Il se peut encore que l'on soit obligé de travailler sur des lots différents de matière. Comment faire pour obtenir des effets qui ne soient pas entachés par l'erreur systématique qui existe entre les mesures du matin et de l'après-midi ou entre les deux lots de matière ? Il s'agit, dans tous ces cas, de variations systématiques dont il faut tenir compte. Nous verrons que le blocking est le moyen de lutte contre ces erreurs systématiques.

Il se peut également que de petites erreurs systématiques modifient les résultats par groupes d'essais. On peut utiliser la randomisation pour que ces erreurs systématiques puissent être considérées comme erreurs aléatoires et que l'on puisse appliquer les tests statistiques.

Enfin, les variations peuvent n'être dues qu'à des causes aléatoires.

Nous commencerons par étudier la nature des différentes erreurs. Nous donnerons ensuite des exemples des différentes stratégies que l'on peut adopter en fonction des connaissances que l'on a des perturbations qui peuvent intervenir en cours d'expérimentation.

8.1 Nature des erreurs

Quand on réalise plusieurs fois la même mesure dans les mêmes conditions opératoires, on constate de petites différences entre toutes ces mesures. Ces petites variations forment ce que l'on appelle l'*erreur expérimentale*.

L'origine de ces variations doit être cherchée dans les variations des niveaux de tous les facteurs qui peuvent modifier la réponse. On considère que l'on peut classer les facteurs actifs sur la réponse en deux catégories : les facteurs contrôlés et les facteurs non contrôlés.

Les facteurs contrôlés sont ceux dont l'expérimentateur fixe les niveaux. Il s'agit donc des facteurs pris en compte pour l'exécution du plan et de tous les facteurs fixés à un niveau constant pendant toute l'expérimentation. On considère (hypothèse de la régression linéaire) que ces facteurs n'introduisent pas d'erreurs. On peut illustrer les changements de niveaux subis par un facteur contrôlé (facteur 1) lors de l'exécution d'un plan 2^3 par la figure 8.1.

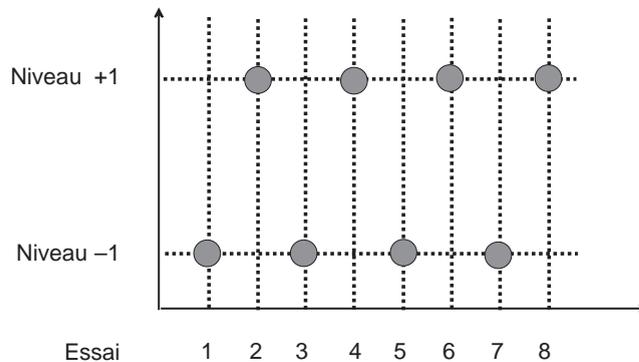


Figure 8.1 – Facteur contrôlé : les changements de niveaux sont imposés par l'expérimentateur.

Les facteurs non contrôlés sont les autres facteurs, ceux dont les niveaux ne sont pas fixés par l'expérimentateur. Ces facteurs peuvent faire varier la réponse dans des limites plus ou moins importantes selon les variations imprévues et inconnues de leurs niveaux. On distingue plusieurs types de facteurs non contrôlés donnant lieu aux trois techniques de lutte suivantes : blocking, plans anti-dérive et randomisation.

8.1.1 Blocking

Considérons les facteurs dont le niveau reste constant pendant une série d'essais. Par exemple, si l'on effectue la première partie d'un plan d'expériences pendant

l'été et la seconde partie en hiver, il y a deux campagnes d'essais. On suppose qu'un facteur non contrôlé conserve le *niveau a* pendant la première campagne et le *niveau b* pendant la seconde campagne. Cependant, les niveaux *a* et *b* sont différents. On peut aussi imaginer que le plan soit exécuté à la même période mais par deux opérateurs différents ou dans deux laboratoires différents. Une application fréquente en agriculture se présente lorsque l'on doit mener les essais sur deux terrains différents. Il peut y avoir un écart constant entre les réponses mesurées à deux époques ou par deux opérateurs ou par deux laboratoires ou sur deux terrains. La figure 8.2 illustre les variations de niveaux d'un facteur non contrôlé lors de l'exécution d'un plan 2^4 en deux campagnes.

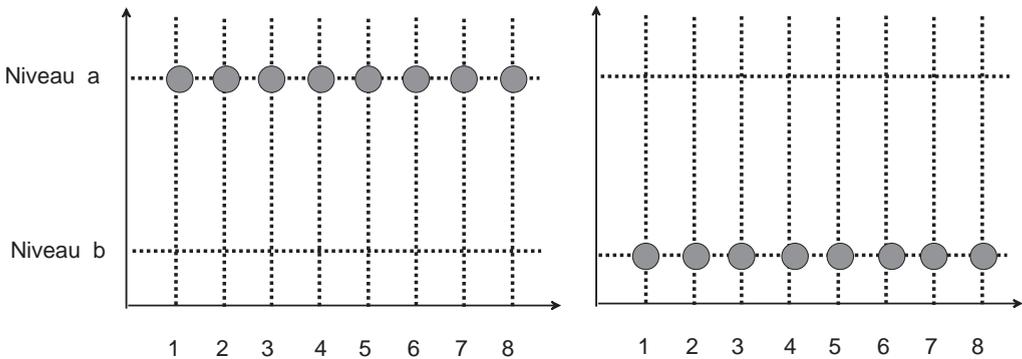


Figure 8.2 – Facteur non contrôlé (*blocking*) : les niveaux de la première campagne sont constants mais différents de ceux de la deuxième.

On trouve souvent ce genre de situation et nous avons la chance que les plans factoriels offrent une méthode pour lutter contre ce type d'erreur systématique. Cette méthode est le *blocking*. Il est possible malgré la perturbation introduite par le facteur non contrôlé, d'obtenir les effets des facteurs étudiés dans le plan sans qu'ils soient faussés par les variations de ce facteur non contrôlé.

Le *blocking* est une technique très employée.

8.1.2 Plans anti-dérive

Le niveau d'un facteur non contrôlé peut diminuer ou augmenter régulièrement en fonction des essais. Les réponses seront modifiées par cette variation de niveau et les effets des facteurs du plan seront faussés si l'on ne prend pas de précautions. Cette variation progressive s'appelle une *dérive* et elle est illustrée par la figure 8.3. Comme exemples, on peut citer l'activité d'un catalyseur qui diminue progressivement, le flux d'un liquide qui se réduit à cause du bouchage progressif d'un tuyau, une solution de soude qui s'altère au fur et à mesure du temps qui passe, ou le vieillissement d'une colonne chromatographique qui sépare de moins en moins bien.

Là aussi, nous avons la chance que les plans factoriels offrent une méthode pour lutter contre la dérive. Il existe des plans anti-dérive qui permettent d'obtenir la valeur des effets des facteurs principaux comme s'il n'y avait pas de dérive.

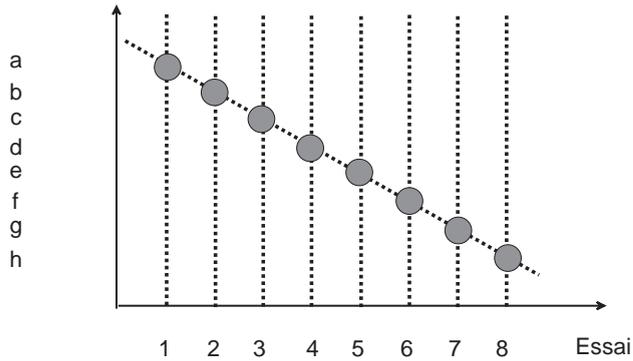


Figure 8.3 – Facteur non contrôlé (dérive) : les changements de niveaux sont progressifs d'un essai à l'autre.

8.1.3 Variations aléatoires

Le niveau d'un facteur peut varier de manière parfaitement aléatoire d'un essai à l'autre (Figure 8.4). Il n'y a pas de moyens de lutter contre les perturbations d'un tel facteur. Il contribue à la formation de l'erreur expérimentale. S'il n'introduit que de toutes petites variations, il ne sera pas trop gênant. S'il introduit de fortes variations de la réponse, il faudra essayer de le contrôler ou d'en réduire l'influence.

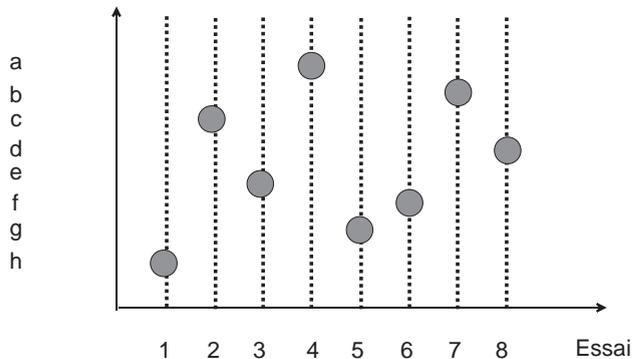


Figure 8.4 – Facteur non contrôlé : les changements de niveaux sont aléatoires.

8.1.4 Petites variations systématiques

Le niveau d'un facteur non contrôlé peut rester stable pendant deux ou trois essais puis changer brusquement pour se stabiliser, de nouveau, pour deux ou trois essais à un autre niveau, et ainsi de suite. Ce facteur introduit des erreurs systématiques qui sont incompatibles avec les tests statistiques. Pour pouvoir appliquer ces tests, on est amené à transformer ces erreurs systématiques en erreurs aléatoires. Pour cela on emploie la *randomisation*.

La randomisation consiste à choisir l'ordre des essais de manière aléatoire. On transforme ainsi la figure 8.5 en figure 8.4. Il est maintenant possible d'utiliser les

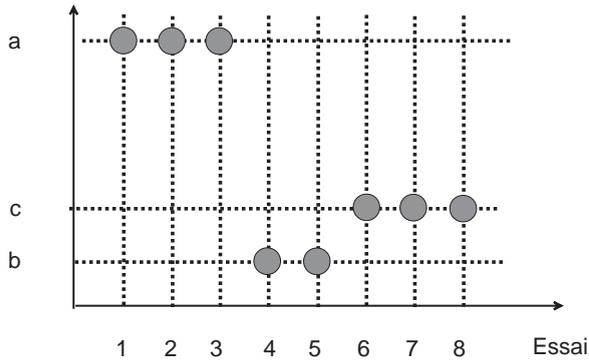


Figure 8.5 – Facteur non contrôlé (randomisation) : les changements de niveaux sont par paquets.

tests statistiques. La transformation des erreurs systématiques en erreurs aléatoires est à la fois une bonne et une mauvaise chose : une bonne chose pour les tests statistiques et une mauvaise car elle augmente la valeur de l'erreur aléatoire. Il devient alors plus difficile de détecter les facteurs peu influents. La procédure correcte est d'éliminer d'abord les erreurs systématiques les plus importantes en utilisant le blocking ou les plans anti-dérive et de randomiser ensuite.

Nous étudierons spécialement le blocking qui est une technique très puissante, les plans anti-dérive et la randomisation.

8.2 Exemple 09 : le *Penicillium chrysogenum* (exemple de blocking)

8.2.1 Préparation du plan d'expériences

■ Description de l'étude

La pénicilline est fabriquée à partir du *Penicillium chrysogenum*. Les responsables de cette fabrication recherchent le milieu nutritif le plus favorable au développement rapide de leur souche. Ils entreprennent un plan d'expériences complet sur les cinq facteurs dont ils veulent connaître l'influence. Ils ont donc 32 essais à réaliser. La durée des essais et les contraintes matérielles ne leur permettent pas de réaliser tous les essais en une seule fois. Ils doivent opérer en deux temps. Ils effectuent une première campagne de 16 essais, font une pause obligatoire de plusieurs semaines puis réalisent une deuxième campagne de 16 essais. Le problème est qu'entre les deux campagnes, beaucoup de facteurs non contrôlés ont pu changer de niveaux : la souche, la température, les opérateurs, etc. Malgré ces changements possibles et imprévisibles, les chercheurs souhaitent obtenir les effets comme si les facteurs non contrôlés n'existaient pas. Ils ont donc recours au blocking.

■ Réponses

La réponse est le poids de pénicilline obtenu pour chaque essai.

■ Objectif de l'étude

Définir la composition de la solution nutritive pour avoir le meilleur rendement en pénicilline.

■ Facteurs

Les facteurs d'étude, c'est-à-dire les facteurs dont les niveaux sont contrôlés au cours du plan d'expériences, sont les concentrations des produits suivants :

- Facteur 1 : liqueur de maïs.
- Facteur 2 : lactose.
- Facteur 3 : précurseur.
- Facteur 4 : nitrate de sodium.
- Facteur 5 : glucose.

■ Domaine d'étude

Le domaine est précisé dans le tableau 8.1.

Tableau 8.1 – Domaine d'étude.

Facteurs	Niveau -1	Niveau +1
Liqueur de maïs (1)	2 %	3 %
Lactose (2)	2 %	3 %
Précurseur (3)	0	0,05 %
Nitrate de sodium (4)	0	0,3 %
Glucose (5)	0	0,5 %

■ Facteur de blocking

En plus des cinq facteurs du plan, il faut prendre en considération le fait que l'on effectue les essais en deux campagnes. On introduit ainsi un facteur supplémentaire : la campagne. Les 16 premiers essais sont au niveau moins du facteur campagne et les 16 essais de la deuxième campagne sont au niveau plus. Ce facteur supplémentaire est le *facteur de blocking*. Le niveau du facteur de blocking reste constant à un niveau donné, mais inconnu, pendant toute la première campagne et reste constant à un autre niveau (lui aussi inconnu) pendant toute la seconde campagne. Il s'agit d'un facteur supplémentaire dont on ne connaît pas la valeur des niveaux.

■ Choix du plan d'expériences

En réalité, on ne réalise pas un plan complet 2^5 mais un plan fractionnaire 2^{6-1} . Il faut donc aliaser le facteur de blocking sur une interaction. Comme il n'y a qu'un facteur supplémentaire, on choisit l'interaction d'ordre le plus élevé :

$$6 = 12345$$

Les essais du plan complets 2^5 vont se partager en deux groupes, celui dont les signes de l'interaction 12345 sont positifs et celui dont les signes de l'interaction 12345 sont négatifs. Bien sûr, c'est avant d'entreprendre les essais qu'il faut faire cette distinction pour attribuer convenablement les essais, soit à la première soit à la deuxième campagne.

■ Confusion

Le générateur d'alias permet grâce au calcul de Box et à la relation d'équivalence de savoir comment les coefficients du modèle sont regroupés dans les alias. En particulier, le facteur de blocking est aliéné avec l'interaction 12345 :

$$\ell_6 = \ell_{12345} = a_6 + a_{12345}$$

L'interaction 12345 risque d'apparaître élevée. S'il y a des interactions entre le facteur de blocking et les facteurs étudiés, on risque également de voir des interactions d'ordre 2 élevées :

$$\ell_{16} = a_{16} + a_{2345} \qquad \ell_{26} = a_{26} + a_{1345} \qquad \text{etc.}$$

8.2.2 Construction du plan et vérification de l'intérêt du blocking

■ Construction du plan

La construction des plans avec blocking est exactement la même que celle des plans fractionnaires. On écrit le plan de base 2^5 et on calcule les signes de l'interaction 12345 (Tableau 8.2). Mais pour les plans avec blocking, on regroupe les essais en fonction du signe de l'interaction. Le premier bloc est celui pour lequel le signe de l'interaction 12345 est moins. Le premier bloc est constitué des 16 essais n° 1, 18, 19, 4, 21, 6, 7, 24, 25, 10, 11, 28, 13, 30, 31 et 16. Le second bloc est celui pour lequel le signe de l'interaction 12345 est plus. Il est constitué des 16 essais n° 17, 2, 3, 20, 5, 22, 23, 8, 9, 26, 27, 12, 29, 14, 15 et 32.

Tableau 8.2 – Construction du plan avec blocking.

Essai n°	1	2	3	4	5	Bloc = 12345
1	-1	-1	-1	-1	-1	-1
18	1	-1	-1	-1	1	-1
19	-1	1	-1	-1	1	-1
4	1	1	-1	-1	-1	-1
21	-1	-1	1	-1	1	-1
6	1	-1	1	-1	-1	-1
7	-1	1	1	-1	-1	-1

Tableau 8.2 – Construction du plan avec blocking. (Suite)

Essai n°	1	2	3	4	5	Bloc = 12345
24	1	1	1	-1	1	-1
25	-1	-1	-1	1	1	-1
10	1	-1	-1	1	-1	-1
11	-1	1	-1	1	-1	-1
28	1	1	-1	1	1	-1
13	-1	-1	1	1	-1	-1
30	1	-1	1	1	1	-1
31	-1	1	1	1	1	-1
16	1	1	1	1	-1	-1
17	-1	-1	-1	-1	1	1
2	1	-1	-1	-1	-1	1
3	-1	1	-1	-1	-1	1
20	1	1	-1	-1	1	1
5	-1	-1	1	-1	-1	1
22	1	-1	1	-1	1	1
23	-1	1	1	-1	1	1
8	1	1	1	-1	-1	1
9	-1	-1	-1	1	-1	1
26	1	-1	-1	1	1	1
27	-1	1	-1	1	1	1
12	1	1	-1	1	-1	1
29	-1	-1	1	1	1	1
14	1	-1	1	1	-1	1
15	-1	1	1	1	-1	1
32	1	1	1	1	1	1

■ Vérification de l'intérêt du blocking

On peut vérifier l'intérêt du blocking sur cet exemple. Si, au cours de la seconde campagne toutes les réponses sont augmentées de 10 points par rapport à la première campagne à cause des facteurs non contrôlés, on peut calculer l'incidence de ce décalage sur les coefficients en attribuant la valeur 10 à tous les essais de la seconde campagne. Le fichier « Penicilline_verif.jmp » a été construit pour que vous puissiez faire cette vérification. Faites le calcul des coefficients en adoptant le modèle avec toutes les interactions jusqu'à l'ordre 5. Vous constaterez que seuls le coefficient constant et l'interaction choisie pour alier le facteur de blocking sont affectés par le décalage entre les deux campagnes. Tous les autres coefficients sont épargnés. Tout se passe, pour eux, comme si l'on avait réalisé tous les essais en une seule fois, l'influence de la campagne a été gommée (Figure 8.6) pour tous les coefficients, excepté le coefficient constant et le coefficient de l'interaction de blocking.

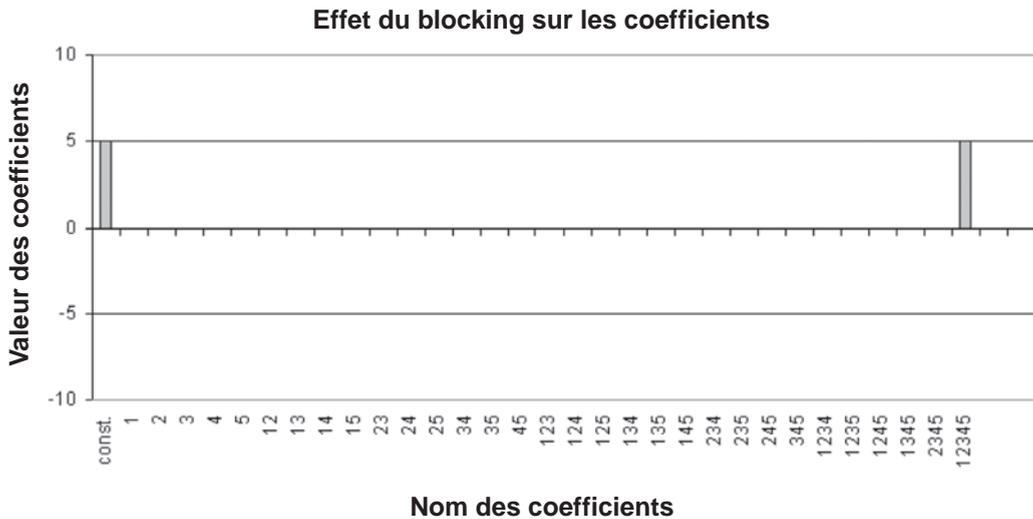


Figure 8.6 – Influence de l'erreur de blocking sur les coefficients.

8.2.3 Réalisation des essais

La première campagne d'essais est réalisée selon les indications du plan d'expériences. Les 16 essais du premier plan sont exécutés dans un ordre aléatoire. Au blocking qui permet de renvoyer l'erreur systématique sur l'interaction 12345, les expérimentateurs ont pris une précaution supplémentaire : une randomisation des essais pour s'affranchir d'éventuelles petites erreurs systématiques. Les essais portent à côté de leur nom un chiffre entre parenthèses. Ce chiffre indique l'ordre d'exécution de l'essai. Par exemple, l'essai n° 18 a été exécuté en 16^e position et il figure à la 2^e ligne du tableau 8.3.

À ce stade il est possible de faire une interprétation des premiers résultats car on peut calculer les 5 effets des facteurs principaux et leurs 10 interactions d'ordre 2. Il a été réalisé la moitié d'un plan complet 2^5 , soit un plan 2^{5-1} avec comme générateur d'aliases $\mathbf{I} = -12345$.

Tableau 8.3 – Première campagne d'essais.

Essai n°	Maïs (1)	Lactose (2)	Précurseur (3)	Nitrate (4)	Glucose (5)	Réponse
1 (5)	-1	-1	-1	-1	-1	142
18 (16)	1	-1	-1	-1	1	106
19 (11)	-1	1	-1	-1	1	88
4 (10)	1	1	-1	-1	-1	109
21 (6)	-1	-1	1	-1	1	113
6 (9)	1	-1	1	-1	-1	162
7 (1)	-1	1	1	-1	-1	200
24 (3)	1	1	1	-1	1	79
25 (7)	-1	-1	-1	1	1	101
10 (2)	1	-1	-1	1	-1	108
11 (13)	-1	1	-1	1	-1	146
28 (15)	1	1	-1	1	1	72
13 (4)	-1	-1	1	1	-1	200
30 (14)	1	-1	1	1	1	83
31 (8)	-1	1	1	1	1	145
16 (12)	1	1	1	1	-1	118

8.2.4 Interprétation des résultats du plan

La deuxième campagne d'essais est réalisée quelques mois plus tard. Le tableau 8.4 indique les résultats des mesures.

Tableau 8.4 – Seconde campagne d'essais.

Essai n°	Maïs (1)	Lactose (2)	Précurseur (3)	Nitrate (4)	Glucose (5)	Réponse
17 (23)	-1	-1	-1	-1	1	106
2 (31)	1	-1	-1	-1	-1	114
3 (18)	-1	1	-1	-1	-1	129
20 (28)	1	1	-1	-1	1	98
5 (25)	-1	-1	1	-1	-1	185
22 (17)	1	-1	1	-1	1	88
23 (30)	-1	1	1	-1	1	166
8 (20)	1	1	1	-1	-1	172
9 (26)	-1	-1	-1	1	-1	148
26 (22)	1	-1	-1	1	1	114
27 (19)	-1	1	-1	1	1	140
12 (32)	1	1	-1	1	-1	95
29 (21)	-1	-1	1	1	1	130
14 (27)	1	-1	1	1	-1	164
15 (24)	-1	1	1	1	-1	215
32 (29)	1	1	1	1	1	110

Pour l'interprétation complète du plan, on réunit les 32 essais et on en fait l'analyse globale. On choisit un modèle avec toutes les interactions. On obtient 32 coefficients.

On voit tout de suite que la liqueur de maïs (1), le précurseur (3) et le glucose (5) sont très influents (Tableau 8.5 et Figure 8.7). Il y a aussi l'interaction 35 qui est forte. Et tout à fait bizarrement des interactions d'ordre élevé : 1234, 1345 et 12345.

Tableau 8.5 – Coefficients du modèle (unités codées).

Effet	Valeur	Effet	Valeur
Constante	129,56		
		123	-1,31
Maïs (1)	-17,56	124	-2,87
Lactose (2)	0,56	125	-1,62
Précurseur (3)	16,06	134	1,75
Nitrate (4)	1	135	-3,25
Glucose (5)	-20,87	145	2,81
		234	-2,62
12	-5,94	235	2,75
13	-6,06	245	2,31
14	-5	345	0,56
15	2,62	1234	4
23	4,44	1235	2,62
24	-1	1245	1,81
25	3	1345	4,19
34	-1	2345	1,06
35	-10,5		
45	2,19	12345	6,31

Effet des facteurs - Pénicilline

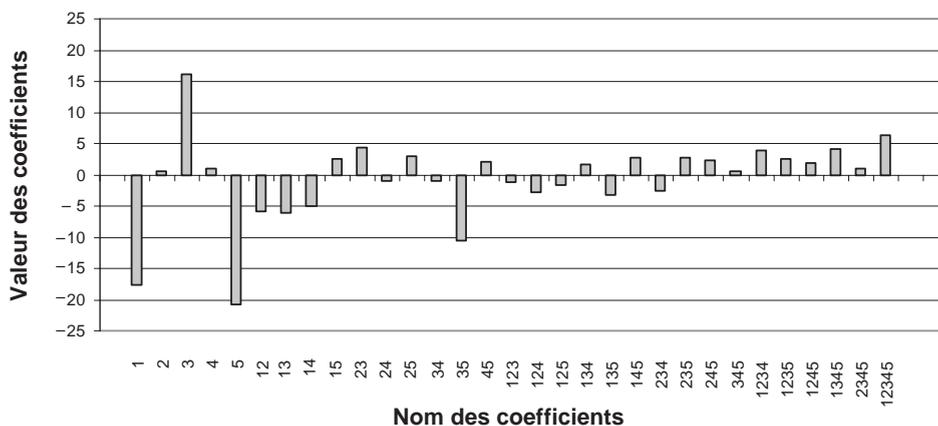


Figure 8.7 – Effets des facteurs principaux et des interactions.

Pour essayer de distinguer les effets à prendre en compte et ceux que l'on peut négliger, on peut utiliser le diagramme de Pareto (Figure 8.8) et le diagramme Daniel (Figure 8.9).

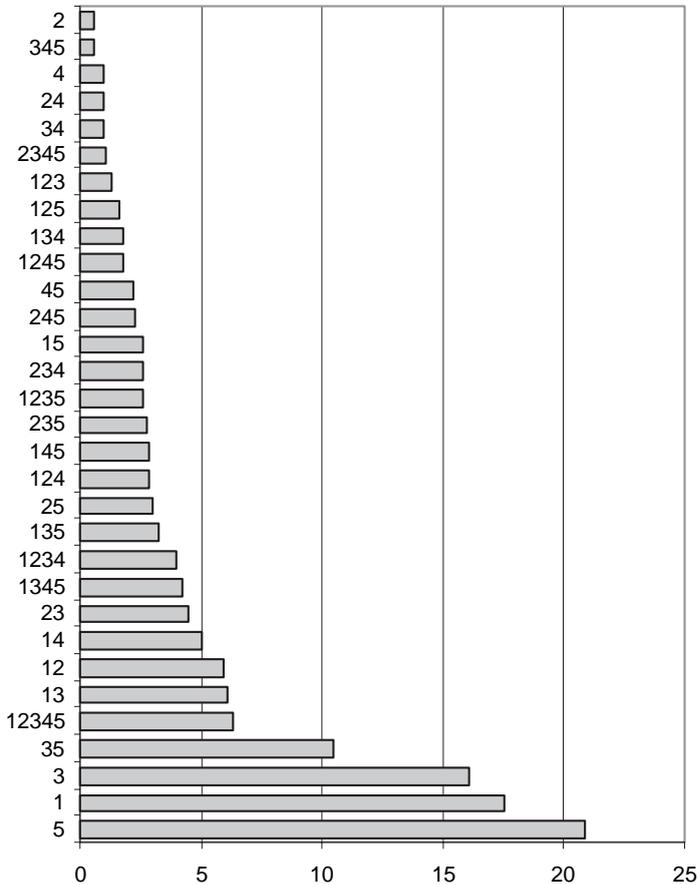


Figure 8.8 – Diagramme de Pareto des effets de l'exemple sur la pénicilline.

L'examen des diagrammes de Pareto et de Daniel montre que l'on peut retenir comme facteurs influents :

- liqueur de maïs (facteur 1),
- précurseur (facteur 3),
- glucose (facteur 5),
- interaction 35.

La coupure entre les coefficients à conserver et les coefficients à rejeter semble se situer aux alentours de la valeur 6. Ce choix peut paraître arbitraire, mais il n'est jamais définitif et peut être modifié pour donner lieu à une nouvelle analyse. Avec cette décision, que faut-il penser de l'interaction 12345 qui est au niveau de la coupure, et qui, selon les hypothèses d'interprétation habituelles, devrait être très faible ?

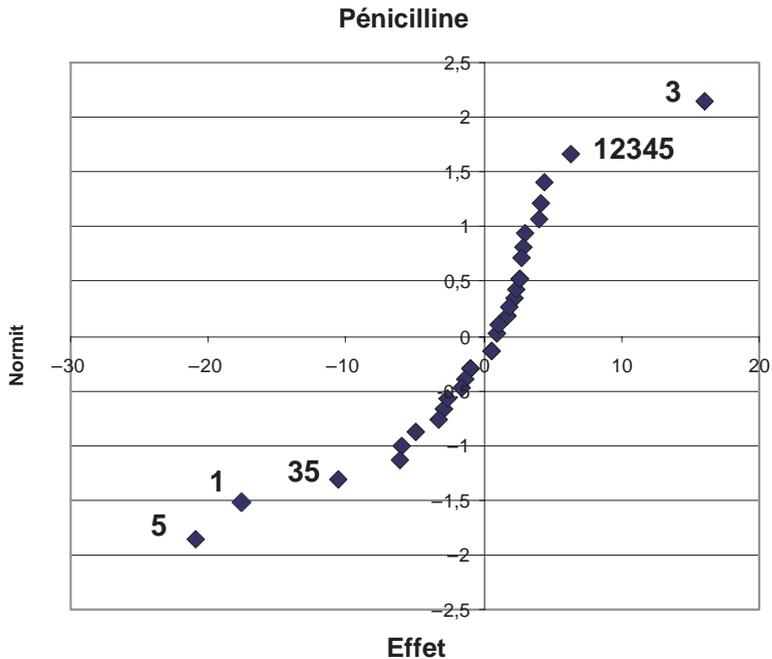


Figure 8.9 – Diagramme de Daniel permettant de distinguer les effets importants.

C'est l'interaction sur laquelle nous avons alié le facteur 6, le facteur de blocking :

$$6 = 12345$$

La relation d'équivalence indique que le facteur de blocking est alié avec l'interaction 12345 :

$$\ell_6 = \ell_{12345} = a_6 + a_{12345}$$

Dans ce contraste, ce n'est pas l'interaction 12345 qui est élevée mais l'effet du facteur 6. Il y a une différence de 12 points entre les réponses de la première campagne et celles de la deuxième. Cette différence n'est pas négligeable et elle aurait fortement faussé les résultats si les expérimentateurs n'avaient pas pris la précaution de faire un blocking.

8.2.5 Conclusion de l'étude

Comment faut-il régler les niveaux des trois facteurs les plus influents pour obtenir la meilleure production possible de pénicilline (Figure 8.10) ?

- La concentration en liqueur de maïs (facteur 1) sera fixée à 2 % (niveau bas).
- La concentration en précurseur (facteur 3) sera fixée à 0,05 % (niveau haut).
- Il n'y aura pas de glucose (niveau bas du facteur 5).

Si on veut encore améliorer légèrement la production, on pourra fixer :

- la concentration en lactose (facteur 2) à 3 % (niveau haut) ;
- la concentration en nitrate de sodium (facteur 4) à 0,3 % (niveau haut).

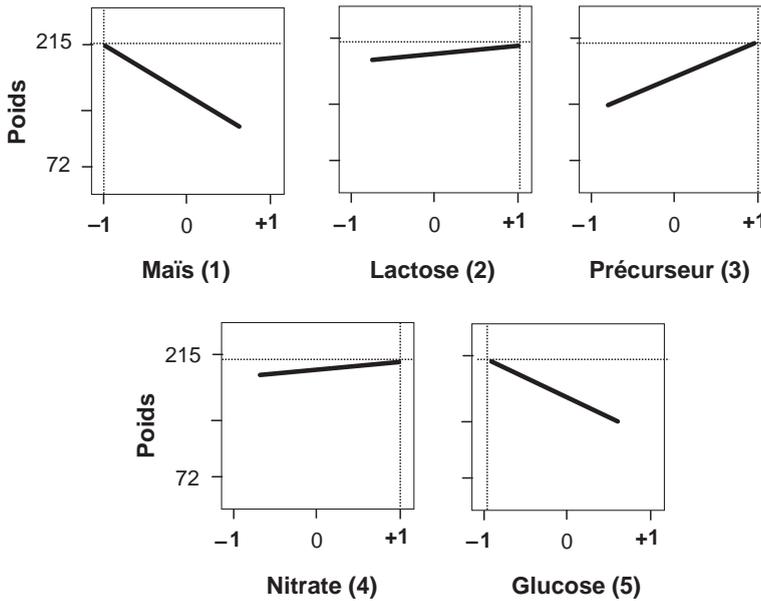


Figure 8.10 – Choix des niveaux des facteurs pour optimiser la production de pénicilline.

8.3 Exemple 10 : les haricots de Yates

8.3.1 Préparation du plan d'expériences

■ Description de l'étude

Yates travaillait avec Fisher à la station expérimentale de Rothamsted en Angleterre dans le Hertfordshire. En 1935, il lança une recherche sur l'influence des engrais sur le rendement d'une espèce de haricots, dont les résultats furent publiés en 1937 ; il s'agit d'un des plus anciens plans d'expériences et il est intéressant de voir que, dès cette époque, les chercheurs avaient su exploiter au mieux les avantages des plans d'expériences dans le milieu agronomique et en particulier utiliser la technique du blocking pour s'affranchir des multiples facteurs non contrôlés que l'on rencontre dans leur domaine. Les expériences étant longues et les terrains d'expérimentation étant très grands il n'est pas nécessaire de restreindre le nombre d'essais et d'envisager des plans fractionnaires. Les précautions à prendre sont d'un autre ordre. En effet, on n'est jamais sûr que deux terrains aient la même fertilité. Sur l'un d'eux, les plantes vont pousser avec exubérance et, sur l'autre, elles seront malingres et rachitiques. Comment obtenir l'influence d'un engrais malgré la différence de fertilité de terrains sur lesquels on mène une même expérimentation ? Il faut d'abord rechercher des terrains assez proches au point de vue fertilité et envisager ensuite un blocking pour renvoyer l'influence de la fertilité sur le facteur de blocking.

Yates voulait étudier quatre engrais et il avait sélectionné quatre terrains pour ses cultures. Comment organiser les plantations pour qu'à la fin des expériences, on

puisse avoir une vue à peu près juste de l'influence de chaque engrais testé ? Il n'a pas choisi la méthode maladroite d'étudier un engrais par terrain. Il a au contraire savamment disposé les essais d'un plan d'expériences sur les quatre terrains et a tenu compte de la différence de fertilité.

■ Réponses

La réponse est le poids de haricots récoltés par essai.

■ Objectif de l'étude

Trouver l'influence des engrais seuls ou associés entre eux.

■ Facteurs

Les facteurs retenus par Yates étaient :

- Facteur 1 : espace entre les rangées de haricots.
- Facteur 2 : quantité de fumure.
- Facteur 3 : quantité de nitrate.
- Facteur 4 : quantité de superphosphate.
- Facteur 5 : quantité de potasse.

■ Domaine d'étude

Le domaine d'étude est précisé dans le tableau 8.6.

Tableau 8.6 – Domaine d'étude.

Facteurs	Niveau -1	Niveau +1
Espace entre rangées (1)	18 inches	24 inches
Fumure (2)	0 tonne à l'acre	10 tonnes à l'acre
Nitrate (3)	0 livre à l'acre	50 livres à l'acre
Superphosphate (4)	0 livre à l'acre	60 livres à l'acre
Potasse (5)	0 livre à l'acre	100 livres à l'acre

■ Choix du plan d'expériences

Comme il y a 5 facteurs qui prennent 2 niveaux, Yates décida de réaliser un plan complet 2^5 . Mais il prit la précaution de faire un blocking pour renvoyer les différences de fertilité sur des interactions et obtenir ainsi la vraie valeur des effets des 5 facteurs du plan.

Il faut répartir les 32 essais sur les 4 terrains dont les fertilités peuvent être différentes. On s'arrange pour mettre 8 essais par terrain. Pour diviser 32 par 4, il faut introduire 2 facteurs de blocking, le facteur 6 et le facteur 7. En effet, la colonne

du facteur 6 contient autant de signes moins que de signes plus. Elle permet de diviser les essais en deux blocs. Il en est de même de la colonne 7. En associant la colonne 6 et la colonne 7, on obtient les 4 blocs (+ +), (+ -), (- +) et (- -), contenant chacun 8 essais.

Tout se passe comme si Yates ne réalisait pas un plan complet 2^5 , mais un plan fractionnaire 2^{7-2} . Comme il y a deux facteurs de blocking, il faut choisir deux interactions. Yates aliase le facteur de blocking 6 sur l'interaction 124 et le facteur de blocking 7 sur l'interaction 135 :

$$6 = 124$$

$$7 = 135$$

Ces deux générateurs indépendants engendrent le GGA :

$$I = 1246 = 1357 = 234567$$

On pourrait calculer à la main la répartition des essais entre les quatre blocs et c'est ce qu'a fait Yates car il n'avait pas d'ordinateur. Il a écrit tous les signes des interactions 124 et 135 et les a groupés selon le tableau 8.7. Le bloc 1 contient tous les essais dont l'interaction 124 possède un signe moins et l'interaction 135 possède un signe moins. Le bloc 1 est composé des essais n° 1, 8, 11, 14, 20, 21, 26 et 31. Ce bloc est attribué à l'un des quatre terrains. On opère de même pour les autres blocs. Aujourd'hui on utilise un logiciel pour faire ce travail, ce qui est beaucoup plus rapide.

Tableau 8.7 – Définition des niveaux des facteurs de blocking pour chaque terrain.

Terrain	124	135
Bloc 1	-1	-1
Bloc 2	+1	-1
Bloc 3	-1	+1
Bloc 4	+1	+1

■ Vérification de l'intérêt du blocking

On utilise le modèle d'un plan complet 2^5 avec toutes les interactions, mais certaines seront faussées par la présence des facteurs de blocking.

Vérifions qu'une différence de fertilité entre les terrains ne modifie pas les effets des facteurs étudiés. Pour cela on suppose que la fertilité du premier bloc décale les réponses de 5, que la fertilité du second bloc décale les réponses de 10, que la fertilité du troisième bloc décale les réponses de 15 et que la fertilité du quatrième bloc décale les réponses de 20.

Seuls le coefficient constant et les interactions choisies pour alier les facteurs de blocking sont affectés par le décalage entre les fertilités des quatre terrains (Figure 8.11). Les autres coefficients ne sont pas affectés par les variations de fertilité. Tout se passe comme si Yates n'avait eu qu'un seul terrain de même fertilité.

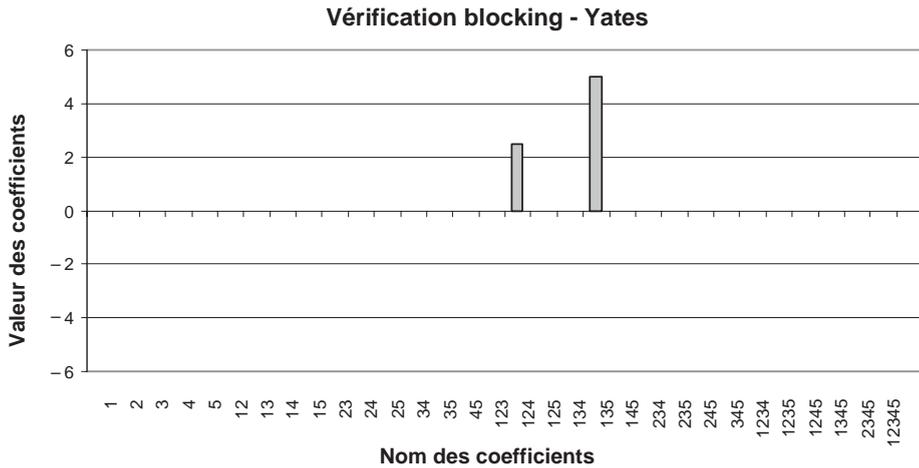


Figure 8.11 – Vérification de l'incidence du blocking.

■ Mise en place des essais

On divise le premier terrain en 8 parcelles et on lui attribue au hasard l'un des quatre blocs. On y dispose les 8 essais correspondant au bloc. On opère de même pour les trois autres terrains.

Comment faut-il disposer les essais sur chaque terrain ? Il est probable qu'il y a, là aussi, de petites variations de fertilité. Il est aussi possible que cette variation ne soit pas aléatoire. La fertilité peut diminuer progressivement quand on se déplace du nord au sud ou d'est en ouest. Toutes les autres variations sont également possibles. Dans ces conditions, si on dispose les essais par ordre de numéro croissant du nord au sud, on aura une corrélation entre l'erreur introduite par la variation de fertilité et le numéro d'ordre des essais. Cette situation est très gênante pour appliquer les tests statistiques qui supposent que les écarts (variation de fertilité) sont indépendants, c'est-à-dire non corrélés entre eux. Pour pouvoir appliquer les tests statistiques, on a l'habitude de répartir les essais au hasard à l'intérieur des blocs, donc ici sur le terrain. C'est ce que fit Yates.

8.3.2 Expérimentation

Après cette phase de préparation, les cultures sont lancées et surveillées jusqu'à la récolte des haricots. Les quantités recueillies sont pesées et le poids de haricots de chaque essai est enregistré. Ces résultats figurent dans le tableau 8.8.

Tableau 8.8 – Plan d'expériences et résultats expérimentaux.

Essai n°	1	2	3	4	5	6 = 124	7 = 135	Bloc	Réponses
1	-1	-1	-1	-1	-1	-1	-1	1	66,5
2	1	-1	-1	-1	-1	1	1	4	36,2
3	-1	1	-1	-1	-1	1	-1	2	74,8

Tableau 8.8 – Plan d'expériences et résultats expérimentaux. (Suite)

Essai n°	1	2	3	4	5	6 = 124	7 = 135	Bloc	Réponses
4	1	1	-1	-1	-1	-1	1	3	54,7
5	-1	-1	1	-1	-1	-1	1	3	68
6	1	-1	1	-1	-1	1	-1	2	23,3
7	-1	1	1	-1	-1	1	1	4	67,3
8	1	1	1	-1	-1	-1	-1	1	70,5
9	-1	-1	-1	1	-1	1	-1	2	56,7
10	1	-1	-1	1	-1	-1	1	3	29,9
11	-1	1	-1	1	-1	-1	-1	1	76,7
12	1	1	-1	1	-1	1	1	4	49,8
13	-1	-1	1	1	-1	1	1	4	36,3
14	1	-1	1	1	-1	-1	-1	1	45,7
15	-1	1	1	1	-1	-1	1	3	60,8
16	1	1	1	1	-1	1	-1	2	64,6
17	-1	-1	-1	-1	1	-1	1	3	63,6
18	1	-1	-1	-1	1	1	-1	2	39,3
19	-1	1	-1	-1	1	1	1	4	51,3
20	1	1	-1	-1	1	-1	-1	1	73,3
21	-1	-1	1	-1	1	-1	-1	1	71,2
22	1	-1	1	-1	1	1	1	4	60,5
23	-1	1	1	-1	1	1	-1	2	73,7
24	1	1	1	-1	1	-1	1	3	92,5
25	-1	-1	-1	1	1	1	1	4	49,6
26	1	-1	-1	1	1	-1	-1	1	74,3
27	-1	1	-1	1	1	-1	1	3	63,6
28	1	1	-1	1	1	1	-1	2	56,3
29	-1	-1	1	1	1	1	-1	2	48
30	1	-1	1	1	1	-1	1	3	47,9
31	-1	1	1	1	1	-1	-1	1	77
32	1	1	1	1	1	1	1	4	61,3

8.3.3 Interprétation des résultats du plan

L'analyse de ce plan commence, comme d'habitude, par le calcul des coefficients du modèle polynomial dont toutes les interactions ont été conservées dans le modèle (Tableau 8.9).

Tableau 8.9 – Coefficients du modèle (unités codées).

Effet	Valeur	Effet	Valeur
Constante	58,91		
		123	0,99
Espace entre rangées (1)	-3,90	124	-5,85
Fumure (2)	7,85	125	-0,76
Nitrate (3)	1,62	134	0,45
Superphosphate (4)	-2,75	135	-3,08
Potasse (5)	3,8	145	-1,74
		234	0,54
12	2,52	235	1,12
13	1,66	245	-0,87
14	1,47	345	-2,42
15	4,37	1234	-0,31
23	2,57	1235	-1,02
24	-0,24	1245	-1,86
25	-1,94	1345	-3,17
34	-2,58	2345	1,56
35	2,175		
45	-0,21	12345	2,39

Ce tableau est illustré par le diagramme à barres de la figure 8.12.

On obtient 32 coefficients dont les plus élevés sont les suivants :

- Le facteur 2 (fumure) est de loin le plus efficace.
- L'interaction 124. Il ne faut pas être étonné de trouver une interaction d'ordre 3 si élevée. Il s'agit en réalité d'un des deux facteurs de blocking. Ce résultat prouve que la fertilité était très différente d'un terrain à l'autre. Sans la précaution du

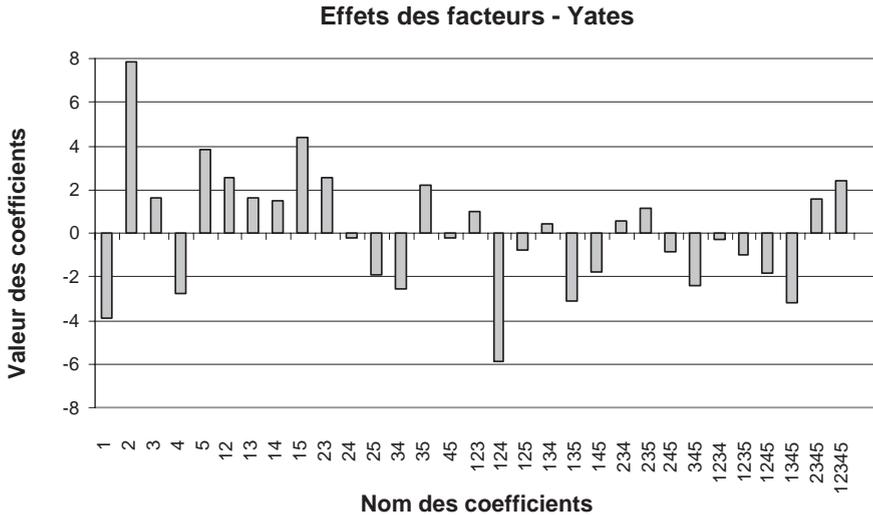


Figure 8.12 – Coefficients du modèle des haricots de Yates.

blocking, l'interprétation du plan aurait été complètement erronée puisqu'on aurait attribué aux facteurs d'étude des rendements provenant des différences de fertilité du terrain.

- L'interaction 15. L'influence de la potasse n'est pas la même selon que les rangs sont serrés ou non.
- Le facteur 1 (espace entre les rangs).
- Le facteur 5 (potasse).
- L'interaction 135. Elle correspond au deuxième facteur de blocking. La forte valeur de cette interaction confirme les fortes disparités de fertilité entre les terrains.
- L'interaction 1345. Il s'agit de l'interaction entre le facteur 4 (superphosphate) et la fertilité des terrains. Cette interaction traduit le fait que cet engrais n'a pas la même influence selon les terrains.

L'influence des autres facteurs et interactions est ensuite de plus en plus faible. Nous retiendrons trois facteurs influents :

- l'espacement entre rangées (facteur 1),
- la quantité de fumure (facteur 2),
- la quantité de potasse (facteur 5).

8.3.4 Conclusion de l'étude

On peut refaire les calculs avec les trois facteurs précédents et leurs interactions d'ordre 2. Les résultats sont reportés sur la figure 8.13.

S'il n'y a ni fumure (niveau bas du facteur 2), ni potasse (niveau bas du facteur 5) et si les rangées sont espacées (niveau haut du facteur 1), le rendement est mauvais (34,5).

On peut améliorer ce rendement de deux manières :

- On resserre les rangées et on ajoute de la fumure. Le rendement est doublé (70,6).

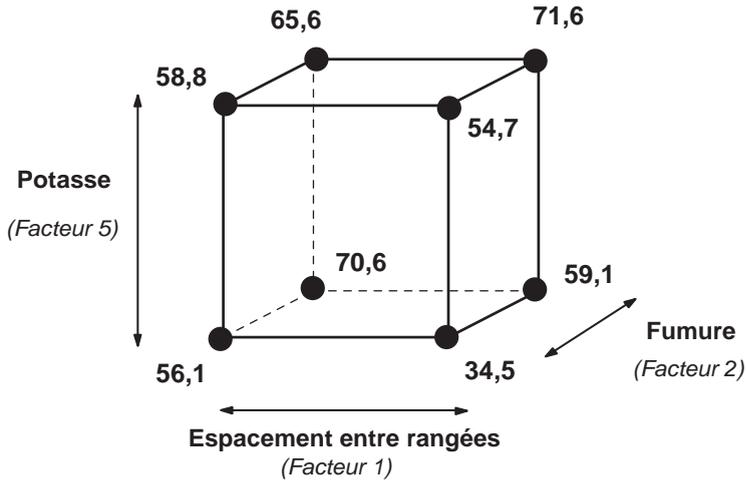


Figure 8.13 – Rendement des haricots en fonction des trois facteurs les plus influents.

- On conserve les rangées espacées, mais on ajoute de la fumure et de la potasse. Là aussi, le rendement est doublé (71,6).

Les deux réglages conduisent à peu près au même rendement. On peut donc choisir le plus économique ou le plus pratique.

8.4 Exemple 11 : le broyeur (exemple de plan anti-dérive)

8.4.1 Introduction

Nous venons de voir un exemple de variations systématiques, les variations par blocs. Il existe aussi des variations systématiques par dérive de la réponse. Il y a dérive lorsque la réponse de chaque expérience est augmentée (ou diminuée) d'une quantité croissante par rapport à la réponse sans dérive. Par exemple, l'usure progressive d'une pièce mécanique ou le vieillissement d'un matériau peut entraîner une variation systématique de la réponse au cours de l'expérimentation. Deux types de dérives seront examinés, la dérive linéaire et la dérive quelconque. Il faut penser à se prémunir contre une dérive.

■ Dérive linéaire

La réponse du premier essai est y_1 quand il n'y a pas de dérive. Quand il y a dérive, elle vaut y'_1 telle que $y'_1 = + y_1 + h$.

Pour le second essai, la réponse sans dérive est y_2 . La réponse avec dérive y'_2 est telle que $y'_2 = + y_2 + 2h$.

Le tableau 8.10 résume les données du problème pour un plan 2^3 , en supposant que l'ordre des essais soit celui correspondant à l'ordre habituel, c'est-à-dire à l'ordre de Yates.

Tableau 8.10 – Dérive linéaire.

Essai	Réponses sans dérive	Réponses avec dérive
1	y_1	$y'_1 = y_1 + h$
2	y_2	$y'_2 = y_2 + 2h$
3	y_3	$y'_3 = y_3 + 3h$
4	y_4	$y'_4 = y_4 + 4h$
5	y_5	$y'_5 = y_5 + 5h$
6	y_6	$y'_6 = y_6 + 6h$
7	y_7	$y'_7 = y_7 + 7h$
8	y_8	$y'_8 = y_8 + 8h$

La matrice des effets d'un plan 2^3 permet de calculer l'incidence de la dérive sur chacun des effets et sur chacune des interactions. Le tableau 8.11 illustre ce calcul.

Tableau 8.11 – Influence de la dérive sur les effets d'un plan 2^3 .

Essai n°	I	1	2	3	12	13	23	123	Réponses
1 (1)	+1	-1	-1	-1	+1	+1	+1	-1	1 h
2 (2)	+1	+1	-1	-1	-1	-1	+1	+1	2 h
3 (3)	+1	-1	+1	-1	-1	+1	-1	+1	3 h
4 (4)	+1	+1	+1	-1	+1	-1	-1	-1	4 h
5 (5)	+1	-1	-1	+1	+1	-1	-1	+1	5 h
6 (6)	+1	+1	-1	+1	-1	+1	-1	-1	6 h
7 (7)	+1	-1	+1	+1	-1	-1	+1	-1	7 h
8 (8)	+1	+1	+1	+1	+1	+1	+1	+1	8 h
Influence dérive	36 h	4 h	8 h	16 h	0	0	0	0	

On en conclut que les effets sont modifiés ainsi :

$$E'_1 = E_1 + 0,5 h$$

$$E'_2 = E_2 + 1 h$$

$$E'_3 = E_3 + 2 h$$

$$E'_{12} = E_{12}$$

$$E'_{13} = E_{13}$$

$$E'_{23} = E_{23}$$

$$E'_{123} = E_{123}$$

$$I' = I + 4,5 h$$

Les quatre interactions ne sont pas influencées par la dérive alors que les effets principaux le sont. Ne peut-on pas s'arranger pour que ce soit l'inverse ? Il suffit de remarquer que c'est l'ordre des signes plus et moins des interactions qui annule l'influence de la dérive. Il faut donc organiser les essais pour que les effets principaux utilisent l'ordre des signes des interactions. Le facteur 1 sera par exemple étudié sur la colonne des signes de l'interaction 123, le facteur 2 sur celle de 12 et le facteur 3 sur celle de 23.

Ce qui conduit à la nouvelle matrice d'expériences où figurent les mêmes essais que dans la matrice initiale mais dans un ordre différent. Avec les hypothèses faites, il faut adopter l'ordre : 7 6 2 3 4 1 5 8. Les effets principaux sont alors indépendants de la dérive, mais les interactions sont évidemment entachées d'une erreur comme le montre le tableau 8.12.

Tableau 8.12 – Influence de la dérive sur les effets d'un plan 2^3 réordonné.

Essai n°	1	1'	2'	3'	1'2'	1'3'	2'3'	1'2'3'	Réponses
7 (1)	+1	-1	-1	-1	+1	+1	+1	-1	1 h
6 (2)	+1	+1	-1	-1	-1	-1	+1	+1	2 h
2 (3)	+1	-1	+1	-1	-1	+1	-1	+1	3 h
3 (4)	+1	+1	+1	-1	+1	-1	-1	-1	4 h
4 (5)	+1	-1	-1	+1	+1	-1	-1	+1	5 h
1 (6)	+1	+1	-1	+1	-1	+1	-1	-1	6 h
5 (7)	+1	-1	+1	+1	-1	-1	+1	-1	7 h
8 (8)	+1	+1	+1	+1	+1	+1	+1	+1	8 h
Influence dérive	36 h	0	0	0	16 h	4 h	0	8 h	

Il faut retenir que lorsqu'il y a dérive, l'ordre des essais n'est pas indifférent sur le résultat et qu'il faut soigneusement choisir cet ordre pour aboutir à des conclusions pertinentes sur les effets principaux et les interactions.

En particulier, on se méfiera des interactions qui, dans ce cas, sont faussées par l'erreur de dérive. On a là un moyen de détecter une dérive car il y a trois valeurs qui sont dans les proportions 1, 2, 4.

L'ordre des essais que nous venons d'indiquer n'est pas le seul qui permette d'obtenir des effets principaux vierges de l'influence d'une dérive linéaire. On peut écrire 144 ordres différents qui réalisent ce même objectif pour un plan 2^3 . L'annexe 3 les indique tous.

Si l'on a pris la précaution d'ordonner les essais suivant les données de ce tableau, il est possible de déceler une dérive (examen des interactions) tout en obtenant de bonnes estimations des effets principaux.

■ Dérive quelconque

L'expérimentateur redoute une dérive, mais il n'a aucune raison de penser qu'elle est linéaire. Il désire mieux la connaître et pouvoir le cas échéant calculer les effets comme s'il n'y avait pas eu de perturbation.

8.4.2 Préparation du plan d'expériences

■ Description de l'étude

Un expérimentateur désire augmenter la quantité de poudre qu'il produit avec un broyeur. Il redoute une usure des mâchoires qui risque d'introduire une dérive et donc d'obtenir des effets faussés.

■ Objectif de l'étude

Obtenir les effets des facteurs principaux comme s'il n'y avait pas de dérive.

■ Réponses

La réponse est la masse de poudre (en grammes) qui sort, à la bonne granulométrie, du broyeur, chaque essai ayant la même durée.

■ Facteurs

Les facteurs à prendre en compte sont :

- la vitesse de rotation (facteur 1),
- la pression d'écrasement (facteur 2),
- la hauteur de l'entrefer (facteur 3).

■ Choix du plan d'expériences

Comme il y a trois facteurs à étudier, le plus simple est de choisir un plan factoriel complet 2^3 . Mais la présence d'une dérive éventuelle incite l'expérimentateur à choisir un plan anti-dérive. Même si la dérive n'est pas linéaire, les effets principaux seront peu affectés par cette cause d'erreur. Mais il vaut mieux, en plus, pouvoir évaluer exactement l'influence de cette dérive. Pour cela il faut insérer des points de répétition dans le plan d'expériences afin de suivre l'évolution de la dérive et de pouvoir apporter les corrections nécessaires à l'obtention des effets comme s'il n'y avait pas de dérive. Le nombre de répétitions dépend de l'enjeu de l'étude et du coût des essais. Il faut prévoir, au moins, un point de répétition au début, au milieu et à la fin du plan. On peut prévoir au plus un point de répétition entre chaque essai du plan. En fonction de l'enjeu de l'étude et du budget l'expérimentateur choisira entre 3 et 9 points au centre pour suivre la dérive. L'exemple est développé avec 9 points au centre.

L'expérimentateur tire donc au hasard l'ordre d'un plan anti-dérive dans la liste des 144 ordres possibles et incorpore un point de répétition entre chaque essai du plan.

L'ordre tiré au hasard est 6 7 3 2 4 1 5 8. Les points au centre sont numérotés avec un 0 devant le chiffre de l'essai. Le plan anti-dérive complet comprend donc $9 + 8 = 17$ essais. Le tableau 8.13 rassemble les expériences réalisées et les réponses obtenues.

8.4.3 Réalisation des essais

Les essais sont scrupuleusement réalisés dans l'ordre prévu par le plan d'expériences et les résultats sont consignés dans le tableau 8.13.

Tableau 8.13 – Matrice d'expériences et résultats bruts des mesures.

Essai n°	Vitesse rotation (1)	Pression (2)	Entrefer (3)	Réponses
01	0	0	0	500
6 (1)	+	–	+	493
02	0	0	0	460
7 (2)	–	+	+	410
03	0	0	0	426
3 (3)	–	+	–	402
04	0	0	0	393
2 (4)	+	–	–	410
05	0	0	0	363
4 (5)	+	+	–	434
06	0	0	0	335
1 (6)	–	–	–	261
07	0	0	0	309
5 (7)	–	–	+	216
08	0	0	0	285
8 (8)	+	+	+	338
09	0	0	0	263
Niveau –	40	4	0,20	
Niveau 0	50	6	0,25	
Niveau +	60	8	0,30	

8.4.4 Interprétation des résultats du plan

L'examen de la suite des valeurs obtenues pour le point au centre montre clairement qu'il y a dérive. On peut la tracer sur un graphique en portant en ordonnées les valeurs de la réponse et en abscisses les numéros des expériences (Figure 8.14).

On portera également sur le même graphique les réponses correspondant aux points expérimentaux du plan 2^3 .

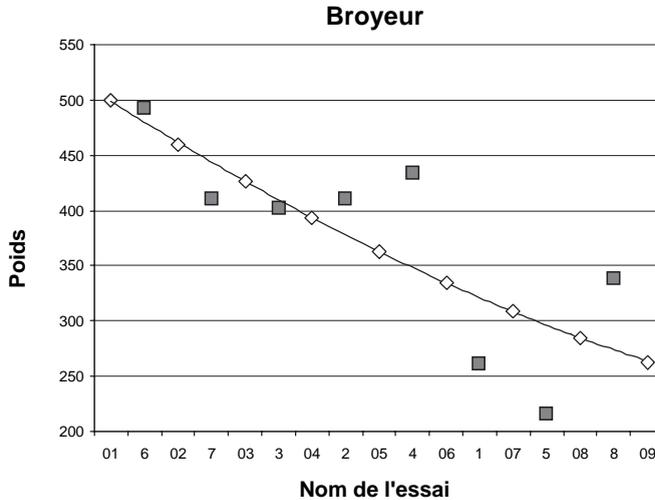


Figure 8.14 – Les répétitions au point central permettent de faire apparaître la dérive.

Bien que la dérive ne soit pas linéaire, le calcul des effets (Tableau 8.14) fait apparaître des interactions dans les rapports proches de 0, 1, 2 et 4. Les effets principaux ne devraient pas être trop loin des valeurs réelles. Il est bien apparent sur cet exemple que les interactions ne sont, en aucun cas, des estimations de l'erreur sur les coefficients. Seule l'interaction 13 pourrait être considérée comme telle. (Rappelons que le calcul des effets et des interactions est effectué sans les points de contrôle au centre.)

Tableau 8.14 – Coefficients obtenus sans corrections (unités codées).

Effet	Valeur
Constante	370,5
Vitesse rotation (1)	48,25
Pression (2)	25,5
Entrefer (3)	-6,25
12	-58,25
13	3
23	-15,75
123	-29

La dérive n'étant pas tout à fait linéaire, il n'est pas possible d'utiliser les formules du paragraphe précédent pour faire les corrections. Nous allons adopter une méthode graphique. Entre les essais n° 01 et 02, l'écart est de 40 g soit 20 g pour l'essai n° 6 exécuté entre les essais n° 01 et 02. On suppose que s'il n'y avait pas eu de dérive, le résultat trouvé aurait été plus fort de 20 g, soit $493 + 20 = 513$ g. Pour l'essai n° 7 (exécuté en second), la valeur moyenne au point milieu a été $1/2(460 + 426) = 443$ g au lieu de 500 g, soit 57 g en moins. Le résultat de l'essai n° 7 doit être augmenté d'autant : $410 + 57 = 467$ g (Tableau 8.15).

Tableau 8.15 – Réponses corrigées de la dérive (unités codées).

Essai n°	Réponses corrigées (g)
6	513
7	467
3	492
2	530
4	584
1	438
5	420
8	566

On effectue de même la correction de toutes les réponses des essais du plan factoriel. On recalcule ensuite les effets et les interactions avec ces réponses corrigées. On obtient de cette façon, les coefficients comme s'il n'y avait pas eu de dérive. Le tableau 8.15 indique les réponses corrigées de la dérive et le tableau 8.16 indique les effets recalculés avec ces nouvelles valeurs.

Tableau 8.16 – Coefficients corrigés de la dérive (unités codées).

Effet	Valeur
Constante	501,25
Vitesse rotation (1)	47
Pression (2)	26
Entrefer (3)	-9,75
12	0,75
13	1
23	-1
123	0,75

8.4.5 Conclusion de l'étude

Les trois facteurs étudiés sont influents sur le rendement en poudre. Le plus important est la vitesse de rotation qui doit être d'au moins 60 tr/min. La pression du ressort doit être réglée le plus fort possible, ici à 8. L'entrefer doit être à 0,20, niveau bas du réglage.

Il n'y a pas d'interaction entre ces facteurs.

De jour en jour, on constate une forte diminution du rendement due à une usure des mâchoires.

Il sera donc nécessaire de changer la nature des mâchoires pour éviter la diminution de rendement. Il faudra également examiner si une vitesse de rotation plus élevée et une pression plus forte ne pourraient pas améliorer encore le rendement.

8.5 Avantages et dangers de la randomisation

Nous avons déjà signalé que la randomisation consiste à choisir l'ordre des essais de manière aléatoire. Le but essentiel est de pouvoir utiliser les tests statistiques même s'il reste de petites erreurs systématiques. Mais la transformation des erreurs systématiques en erreurs aléatoires n'est pas toujours une bonne chose. En effet, l'erreur systématique vient grossir l'erreur aléatoire. Il devient alors plus difficile de détecter les facteurs peu influents. Si les erreurs systématiques sont importantes, la randomisation peut devenir une technique catastrophique. Les erreurs systématiques doivent être éliminées avant la randomisation.

La procédure correcte est d'éliminer d'abord les erreurs systématiques les plus importantes en utilisant le blocking ou les plans anti-dérive et de randomiser ensuite. Quels auraient été les résultats du plan avec dérive si les expérimentateurs avaient simplement randomisé les essais sans se méfier de la dérive ? On suppose que les essais ont été tirés au hasard (ordre 1 5 8 7 4 6 2 3) et qu'aucun point de répétition n'a été réalisé.

À l'aide des résultats déjà trouvés, on peut recalculer les réponses que l'expérimentateur aurait mesurées. La matrice d'expériences serait présentée selon le tableau 8.17 et l'expérimentateur en aurait déduit les effets qui figurent dans le tableau 8.18. Les effets ainsi calculés sont complètement faux, ainsi que les interactions. Nous n'insisterons pas sur les conséquences que peuvent avoir de telles erreurs de conception et d'organisation des essais.

Tableau 8.17 – Résultats bruts des mesures d'un plan simplement randomisé.

Essai n°	Vitesse de rotation (1)	Pression (2)	Entrefer (3)	Réponses
1 (1)	+	-	+	415
5 (2)	-	+	+	358
8 (3)	-	+	-	474

Tableau 8.17 – Résultats bruts des mesures d'un plan simplement randomisé. (Suite)

Essai n°	Vitesse de rotation (1)	Pression (2)	Entrefer (3)	Réponses
7 (4)	+	-	-	342
4 (5)	+	+	-	433
6 (6)	-	-	-	337
2 (7)	-	-	+	332
3 (8)	+	+	+	259
Niveau -	40	4	0,20	
Niveau 0	50	6	0,25	
Niveau +	60	8	0,30	

Tableau 8.18 – Coefficients d'un plan simplement randomisé (unités codées).

Effet	Valeur
Constante	368,75
Vitesse rotation (1)	25,25
Pression (2)	8,25
Entrefer (3)	9
12	51,25
13	2,5
23	22
123	-13

Afin de comparer les différentes techniques d'organisation de l'ordre des essais, les effets obtenus ont été rassemblés dans le tableau 8.19 et illustrés par la figure 8.15.

(1) Plan anti-dérive. Les effets des facteurs principaux sont obtenus vierges de l'influence de la dérive. Mais les interactions sont fausses. L'expérimentateur se doutait de la dérive, mais il ne pouvait réaliser que huit essais, il a choisi les meilleures conditions d'expérimentation mais il ne devra pas tirer de conclusions sur les interactions.

Tableau 8.19 – Comparaison des trois techniques.

Effets	Plan anti-dérive	Réponses corrigées de la dérive	Randomisation
Constante	370,5	501,25	368,75
Vitesse rotation (1)	48,25	47	25,25
Pression (2)	25,5	26	8,25
Entrefer (3)	-6,25	-9,75	9
12	-58,25	0,75	51,25
13	3	1	2,5
23	-15,75	-1	22
123	-29	0,75	-13

Comparaison des trois méthodes

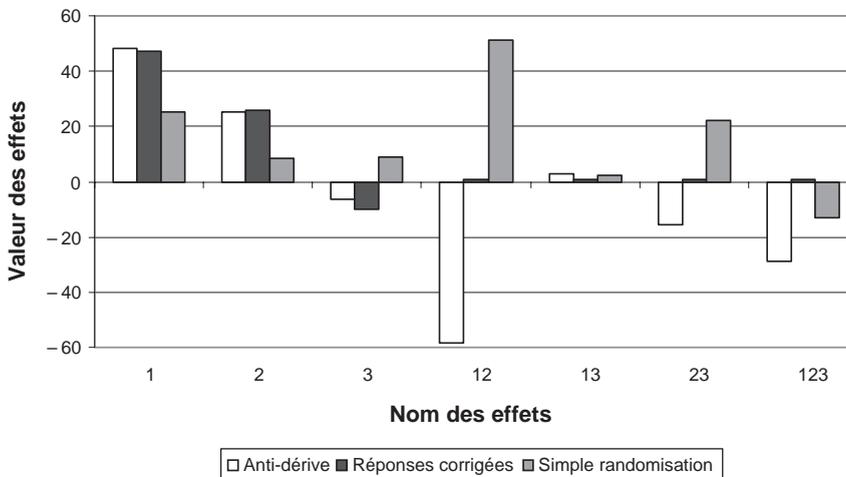


Figure 8.15 – Diagrammes à barres de comparaison des trois méthodes.

(2) **Correction complète de la dérive sur toutes réponses.** Les effets et les interactions sont justes mais au prix d'un nombre élevé d'expériences.

(3) **Simple randomisation sans tenir compte de la dérive.** Les effets des facteurs principaux et des interactions sont faux. Le manque de clairvoyance de l'expérimentateur peut être catastrophique.

Il faut retenir de cet exemple qu'il est nécessaire de s'affranchir d'abord des erreurs systématiques en utilisant le blocking ou les plans anti-dérive, puis de randomiser ensuite.

9 • PLANS POUR SURFACES DE RÉPONSE

Les plans examinés précédemment n'avaient que deux niveaux d'étude par facteur et les modèles mathématiques utilisés étaient du premier degré (avec ou sans interactions) par rapport à chaque facteur. Ces plans sont les plus employés car ils permettent le criblage des facteurs et conduisent parfois à des modélisations simples mais suffisantes. Pourtant, il existe de nombreux cas où il est nécessaire d'avoir une bonne modélisation des phénomènes étudiés et où il faut passer à des modèles mathématiques du second degré. On fait alors appel aux plans pour surfaces de réponse. Ces plans utilisent des modèles polynomiaux du second degré. Nous étudierons les trois plus importants plans de ce type : les plans composites, les plans de Box-Behnken et les plans de Doehlert.

Le modèle mathématique postulé utilisé avec les plans pour surfaces de réponse est un modèle du second degré avec interactions d'ordre 2 :

- pour deux facteurs :

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

- pour trois facteurs :

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

9.1 Présentation des plans composites

Les plans composites se prêtent bien au déroulement séquentiel d'une étude. La première partie de l'étude est un plan factoriel complet ou fractionnaire complété par des points au centre pour vérifier la validité du modèle PDAI (termes du premier degré et termes d'interactions). Si les tests de validation sont positifs (la réponse mesurée au centre du domaine est statistiquement égale à la réponse calculée au même point), l'étude s'achève le plus souvent, mais s'ils sont négatifs, on entreprend des essais supplémentaires pour établir un modèle du second degré. Les essais supplémentaires sont représentés par des points d'expériences situés sur les axes de coordonnées et par de nouveaux points centraux. Les points situés sur les axes de coordonnées sont appelés les *points en étoile*. Les plans composites présentent donc trois parties (Figure 9.1) :

- Le plan factoriel : c'est un plan factoriel complet ou fractionnaire à deux niveaux par facteurs. Les points expérimentaux sont aux sommets du domaine d'étude.

- Le plan en étoile : les points du plan en étoile sont sur les axes et ils sont, en général, tous situés à la même distance du centre du domaine d'étude.
- Les points au centre du domaine d'étude. On prévoit toujours des points expérimentaux situés au centre du domaine d'étude, et cela aussi bien pour les plans factoriels que pour les plans en étoile.

Le nombre total n d'essais à réaliser est la somme des essais du plan factoriel (n_f), des essais du plan en étoile (n_α) et des essais au centre (n_0). Le nombre n des essais d'un plan composite est donné par la relation :

$$n = n_f + n_\alpha + n_0$$

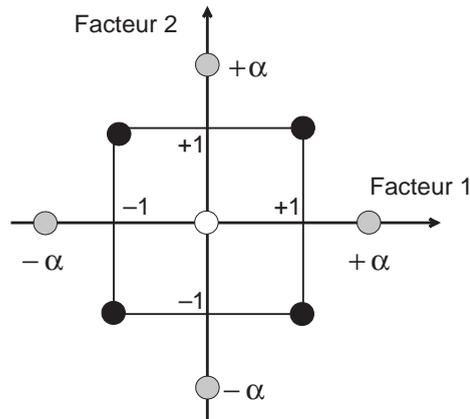


Figure 9.1 – Plan composite pour l'étude de deux facteurs. Les points factoriels sont en noirs, les points en étoile sont en gris clair, les points centraux sont en blanc.

Le nombre de niveaux est 5 pour chacun des facteurs et seulement 3 lorsque l'on a un plan composite à faces centrées.

9.2 Présentation des plans de Box-Behnken

Box et Behnken ont proposé en 1960 ces plans qui permettent d'établir directement des modèles du second degré. Tous les facteurs ont trois niveaux : -1 , 0 et $+1$. Ces plans sont faciles à mettre en œuvre et possèdent la propriété de séquentialité. On peut entreprendre l'étude des k premiers facteurs en se réservant la possibilité d'en ajouter de nouveaux sans perdre les résultats des essais déjà effectués.

Le plan de Box-Behnken pour trois facteurs est construit sur un cube. Pour quatre facteurs ce plan est construit sur un hypercube à quatre dimensions. On place les points expérimentaux non pas aux sommets du cube ou de l'hypercube, mais au milieu des arêtes ou au centre des faces (carrés) ou au centre des cubes. Cette disposition a pour conséquence de répartir tous les points expérimentaux à égale distance du centre du domaine d'étude, donc sur une sphère ou sur une hypersphère suivant le nombre de dimensions. On ajoute des points au centre du domaine d'étude.

Le plan de Box-Behnken pour trois facteurs est illustré par la figure 9.2. Le cube possède 12 arêtes. On a l'habitude d'ajouter des points d'expériences au centre du domaine d'étude, en général trois. Le plan de Box-Behnken pour 3 facteurs possède donc $12 + 3$ essais, soit 15 essais. On pourra remarquer qu'avec 4 points au centre au lieu de 3, on obtient un plan qui répond au critère de presque-orthogonalité.

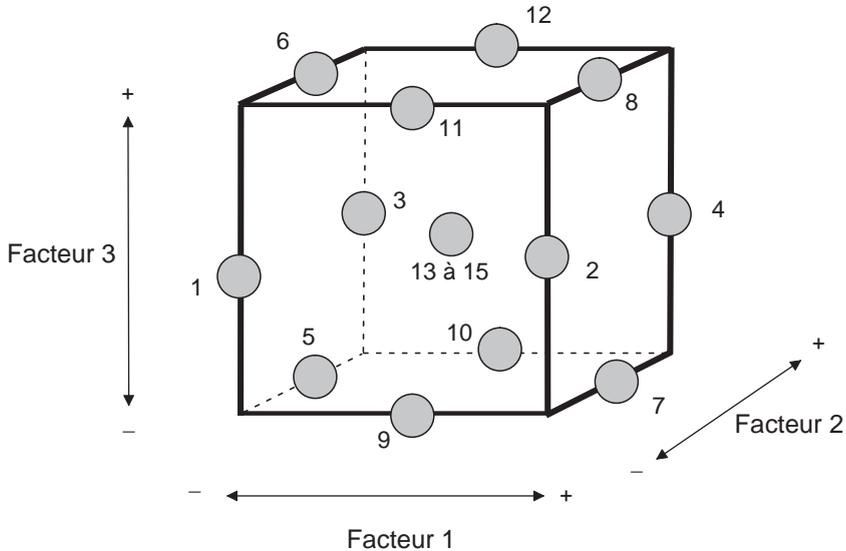


Figure 9.2 – Illustration du plan de Box-Behnken pour trois facteurs.

Il y a douze points d'expériences au milieu des arêtes du cube et trois points au centre.

9.3 Présentation des plans de Doehlert

Les points d'expériences des plans proposés par David H. Doehlert en 1970 remplissent de manière uniforme l'espace expérimental. Pour deux facteurs les points expérimentaux sont situés aux sommets d'un hexagone régulier et il y a un point au centre (Figure 9.3). Ayant sept points expérimentaux, ce plan permet de calculer au moins sept inconnues, donc sept coefficients. Comme les points expérimentaux sont régulièrement répartis dans l'espace expérimental, il sera facile d'étendre le plan vers n'importe quelle direction de l'espace en ajoutant des points qui seront, eux aussi, régulièrement répartis.

Ces plans permettent également l'introduction facile de nouveaux facteurs. Les nouvelles expériences viendront compléter les premières et aucune expérience ne sera perdue. La seule précaution à prendre est de maintenir les facteurs non étudiés à une valeur constante (niveau 0) pendant l'étude des facteurs actifs.

Le tableau 9.1 est la traduction, sous forme de matrice d'expériences, de la figure 9.3. La disposition des points selon la figure 9.3 conduit à cinq niveaux pour le facteur 1 et trois niveaux pour le facteur 2. Avant d'attribuer les numéros aux facteurs, on vérifiera qu'ils sont compatibles avec ces nombres de niveaux.

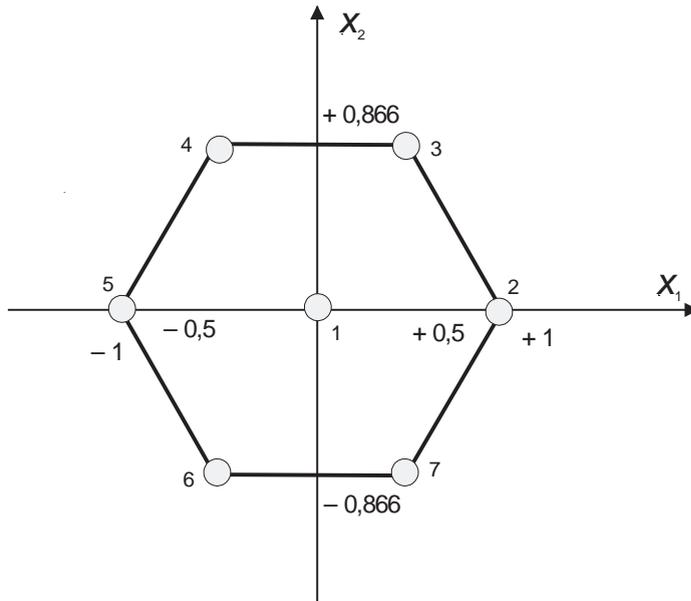


Figure 9.3 – Plan de Doehlert pour l'étude de deux facteurs. Les points sont régulièrement disposés sur un hexagone et il y a un point central.

Tableau 9.1 – Plan de Doehlert pour deux facteurs.

Essai n°	Facteur 1	Facteur 2	$0,866 = \frac{\sqrt{3}}{2}$
1	0	0	
2	+ 1	0	
3	+ 0,5	+ 0,866	
4	- 0,5	+ 0,866	
5	- 1	0	
6	- 0,5	- 0,866	
7	+ 0,5	- 0,866	

Si l'on éprouve quelques difficultés, on peut toujours adopter un autre plan de Doehlert en faisant tourner l'hexagone. La figure 9.4 illustre une autre disposition du même plan après une rotation de 90°. C'est maintenant le facteur 1 qui a 3 niveaux et le facteur 2 qui en a 5.

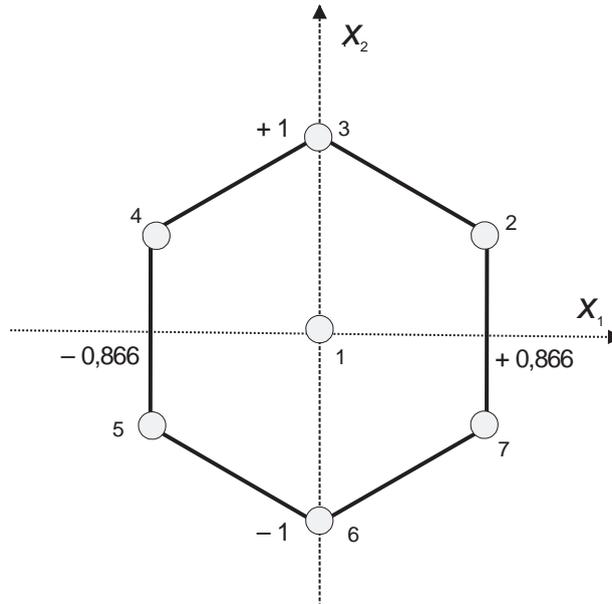


Figure 9.4 – Autre disposition possible des points d'un plan de Doehlert pour l'étude de deux facteurs. Les points sont toujours régulièrement disposés dans l'espace expérimental.

Tous les points du plan de Doehlert sont sur un cercle de rayon unité (en grandeurs centrées réduites). Le domaine défini par les plans de Doehlert est un domaine sphérique, un cercle dans un espace à deux dimensions, une sphère dans un espace à trois dimensions, une hypersphère dans un espace à plus de trois dimensions.

Si les résultats recherchés ne sont pas dans le domaine d'étude, on peut étendre ce domaine dans la direction où l'on a le plus de chances de trouver la solution souhaitée. Il suffit d'ajouter trois points d'expériences (Tableau 9.2) et l'on retrouve un nouveau plan de Doehlert (Figure 9.5). En effet, les points n° 2, 1, 7, 8, 9, 10 et 3 forment un nouvel hexagone. On peut étendre le plan d'expériences dans les autres directions.

Tableau 9.2 – Points d'extension du plan de Doehlert pour deux facteurs.

Essai n°	Facteur 1	Facteur 2
8	+1,5	-0,866
9	+2	0
10	+1,5	+0,866

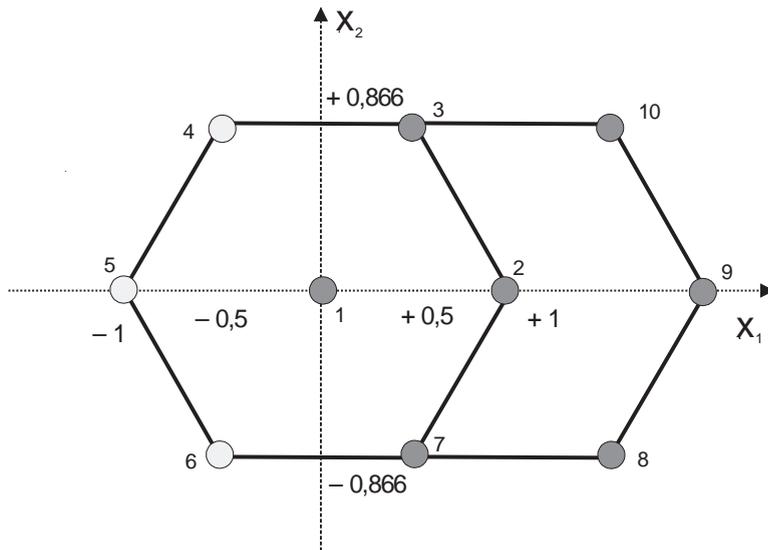


Figure 9.5 – Extension d'un plan de Doehlert à deux facteurs. Trois points suffisent pour retrouver un nouveau plan de Doehlert.

9.4 Exemple 12 : la rectification du contremaître (exemple de plan composite)

9.4.1 Préparation du plan d'expériences

■ Description de l'étude

Un contremaître n'est pas content de l'état de surface des pièces métalliques qui sont fabriquées dans son atelier. Le dernier usinage, celui qui donne l'aspect final à la pièce, est la rectification. C'est donc cette technique que le contremaître cherche à améliorer. L'état de surface des pièces est caractérisé par la rugosité et par le nombre de pics par unité de longueur. Le contremaître serait enchanté s'il pouvait obtenir des pièces de rugosité inférieure à 0,150 et un nombre de pics inférieur à 50. S'il atteint cet objectif, l'état de surface de ses pièces serait considéré comme parfait et ses clients auraient tout lieu d'être satisfaits.

Pour obtenir l'état de surface voulu, le contremaître a réalisé des essais en se fiant à son bon sens et à son expérience. Mais, toutes ses tentatives ont échoué et les résultats semblent même contradictoires. Tantôt une augmentation de la vitesse de coupe améliore la qualité, tantôt elle la détériore. Ne sachant plus que faire, le contremaître décide de mettre en œuvre les recommandations du stage de plans d'expériences qu'il vient de suivre.

La pièce métallique est fermement maintenue sur un support. Une meule arrache une fine pellicule de métal à la surface de la pièce. Cet outil est un disque abrasif qui tourne à grande vitesse et avance lentement en arrachant de petits copeaux de métal (Figure 9.6). L'épaisseur de métal arraché à chaque passe est réglable. La

dernière passe est celle qui contribue le plus à l'aspect final de la pièce. Il faut alors éviter d'exercer de trop fortes contraintes sur la meule et on n'enlève qu'une très fine couche de matière.

Les réglages en usage dans l'atelier sont :

- vitesse d'avancement (0,9 m/min) ;
- vitesse tangentielle de coupe rapide (20 m/s).

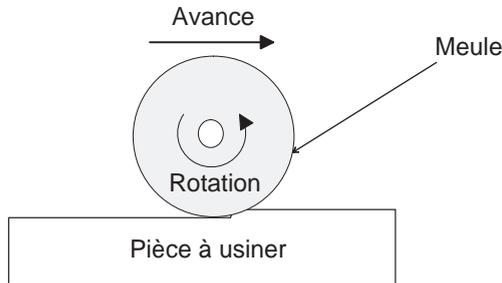


Figure 9.6 – Schéma d'usinage d'une pièce par rectification. La meule avance en tournant et elle aplatit la surface.

■ Réponses et objectifs de l'étude

Les réponses choisies par l'expérimentateur sont la rugosité et le nombre de pics par unité de longueur (Figure 9.7). Les réponses sont définies ainsi :

- **Rugosité** : la rugosité est mesurée par une méthode normalisée. Il faut obtenir la valeur la plus faible possible. Une valeur inférieure à 0,150 serait une réussite. L'écart-type de cette mesure est de 0,002 unité.
- **Pics par unité de longueur** : on compte le nombre de pics par unité de longueur. Ce nombre doit être le plus petit possible. On vise un nombre de pics inférieur à 50 par unité de longueur. L'écart-type de cette mesure est de deux unités.

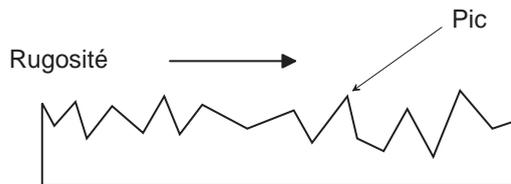


Figure 9.7 – Grossissement de la surface rectifiée. Lors de l'usinage, des pics sont créés par arrachement de matière à la surface de la pièce. Suivant leur nombre et leur hauteur, ces pics sont responsables de la rugosité et de l'aspect de la surface usinée.

■ Facteurs

Le contremaître retient deux facteurs :

- Facteur 1 : vitesse d'avancement de la meule (en m/min).
- Facteur 2 : vitesse tangentielle de coupe (en m/s). Cette vitesse est liée à la vitesse de rotation et au diamètre de la meule.

De nombreux autres facteurs sont fixés : la forme et la matière de la meule, la qualité d'abrasion de la meule, le métal travaillé, la profondeur de la dernière passe (0,02 mm), etc.

■ Domaine d'étude

Les niveaux haut et bas de chaque facteur sont définis comme l'indique le tableau 9.3.

Tableau 9.3 – Domaine d'étude.

Facteur	Niveau -1,21	Niveau -1	Niveau 0	Niveau +1	Niveau +1,21
Vitesse d'avancement (1)	0,74	0,9	1,65	2,4	2,56
Vitesse de coupe (2)	13,95	15	20	25	26,05

■ Choix du plan expérimental

Le contremaître commence par un plan factoriel classique mais il soupçonne qu'il sera obligé de poursuivre l'étude par un plan pour surfaces de réponse. Il prévoit donc deux points de contrôle au centre du domaine d'étude.

9.4.2 Expérimentation (plan factoriel)

Les résultats des essais sont rassemblés dans le tableau 9.4.

Tableau 9.4 – Plan factoriel et résultats.

Essai n°	Avance (1)	Coupe (2)	Rugosité × 1 000	Pics
1	-1	-1	194	77,8
2	+1	-1	282	68,4
3	-1	+1	120	65,3
4	+1	+1	91	96,1
5	0	0	233	63,8
6	0	0	235	61,9
Niveau -1	0,9	15		
Niveau 0	1,65	20		
Niveau 1	2,4	25		

9.4.3 Interprétation des résultats du plan factoriel

Les résultats des calculs figurent dans le tableau 9.5. Rappelons que les coefficients sont calculés avec les quatre essais du plan et sans les points de contrôle.

Tableau 9.5 – Coefficients du modèle (unités codées).

Effet	Rugosité $\times 1\,000$	Pics
Constante	171,75	76,9
Avance (1)	14,75	5,35
Coupe (2)	-66,25	3,8
12	-29,25	10,05

■ Vérification du modèle PDAI

On vérifie la validité du modèle du premier degré avec interactions en comparant la réponse calculée au point de contrôle avec la réponse mesurée en ce même point. Pour la rugosité, la réponse calculée au centre du domaine est égale à la constante du modèle et l'on prend une marge de trois écarts-types :

$$\hat{a}_0 = 171,75 \pm 3 \times \frac{2}{\sqrt{4}} = 171,75 \pm 3$$

Il y a donc 99,74 chances que la moyenne de la population de la réponse calculée soit dans l'intervalle [168,75 ; 174,75].

La réponse mesurée est égale à :

$$y_0 = \frac{233 + 235}{2} = 234 \pm 3 \times \frac{2}{\sqrt{2}} = 234 \pm 4,24$$

Il y a donc 99,74 chances que la moyenne de la population de la réponse mesurée soit dans l'intervalle [229,76 ; 238,24].

Les deux intervalles ne se recouvrent pas (Figure 9.8). On peut conclure avec une très grande probabilité de ne pas se tromper que les deux valeurs ne sont pas égales.

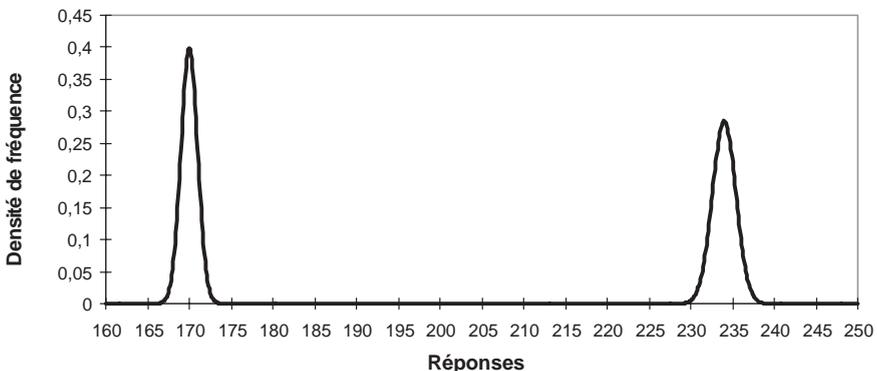


Figure 9.8 – Comparaisons des réponses mesurées et calculées.

On peut aussi employer la formule de comparaison des moyennes (voir annexe 2) :

$$t = \frac{|\hat{a}_0 - y_0|}{\sqrt{\sigma_{a_0}^2 + \sigma_{y_0}^2}} = \frac{171,75 - 234}{\sqrt{1 + 2}} = \frac{62,25}{1,732} = 35,94$$

La différence entre les deux réponses est de 35,94 écarts-types de la différence. La probabilité correspondante est tellement faible qu'elle correspond à la certitude que les deux valeurs ne sont pas égales.

La conclusion est que le modèle PDAI ne convient pas et qu'il faut passer à un modèle du second degré. Pour cela il faut ajouter des points en étoile et des points au centre.

■ Position des points en étoile

C'est un sujet qui a passionné de nombreux statisticiens et il existe des livres entiers sur cette question. En effet, il faut savoir que la précision des coefficients du modèle postulé est influencée par la position des points du plan d'expériences. Selon que les points sont bien ou mal placés, on peut obtenir les coefficients avec une bonne ou une mauvaise précision. Avec les plans à deux niveaux basés sur les matrices d'Hadamard (plans factoriels complets, fractionnaires et plans de Plackett et Burman), on est sûr que les points sont toujours bien placés. Avec les plans du second degré, il n'en est plus de même. Il faut choisir un critère de qualité pour les coefficients. En fonction de ce critère, on calcule la meilleure position possible des points d'expériences. On suppose que les points axiaux sont à la même distance (en grandeurs centrées réduites) du centre du domaine d'étude et cette distance est notée α .

Par exemple, on peut vouloir que l'erreur de prédiction soit la même pour des distances également éloignées du centre du domaine. Dans ce cas, on choisit le critère d'isovariance par rotation. La valeur de α est égale à :

$$\alpha = n_f^{1/4}$$

Si l'on veut que le plan composite satisfasse le critère d'isovariance par rotation il faut placer les points en étoile à une distance α égale à la racine quatrième du nombre de points du plan factoriel.

On peut vouloir que les coefficients répondent au critère de *presque-orthogonalité*, il faut alors choisir α tel que :

$$\alpha = \left(\frac{n_f \left(\sqrt{n_0 + n_f + n_\alpha} - \sqrt{n_f} \right)^2}{4} \right)^{1/4}$$

La valeur de α est fonction du nombre de points au centre, du nombre de points du plan factoriel et du nombre de points du plan en étoile. Le tableau 9.6 permet de choisir la valeur de α pour les cas les plus souvent rencontrés.

Tableau 9.6 – Valeur de α en fonction du nombre de points factoriels (n_f), en étoile (n_α) et centraux (n_0) pour le critère de presque-orthogonalité.

Nombre de facteurs	2	3	4	5	5	6	6
Plan	2^2	2^3	2^4	2^{5-1}	2^5	2^{6-1}	2^6
n_f	4	8	16	16	32	32	64
n_α	4	6	8	10	10	12	12
$n_0 = 1$	1	1,215	1,414	1,547	1,596	1,724	1,761
$n_0 = 2$	1,078	1,287	1,483	1,607	1,662	1,784	1,824
$n_0 = 3$	1,147	1,353	1,547	1,664	1,724	1,841	1,885
$n_0 = 4$	1,210	1,414	1,607	1,719	1,784	1,896	1,943

9.4.4 Plan complémentaire pour le second degré

Le contremaître prévoit de faire deux nouveaux points au centre. Pour que le plan composite à deux facteurs respecte le critère de presque-orthogonalité, il faut choisir $\alpha = 1,21$ puisqu'il y a 4 points factoriels, 4 points en étoile et 4 points au centre. Le plan est exécuté et les résultats figurent dans le tableau 9.7.

Tableau 9.7 – Plan d'expériences complémentaire et résultats.

Essai n°	Avance 1	Coupe 2	Rugosité $\times 1\ 000$	Pics
7	-1,21	0	154	52,3
8	+1,21	0	195	60,4
9	0	-1,21	278	87,0
10	0	+1,21	122	95,7
11	0	0	232	61,5
12	0	0	230	60,5
Niveau -1,21	0,74	13,95		
Niveau 0	1,65	20		
Niveau +1,21	2,56	26,05		

9.4.5 Interprétation des résultats du plan composite

L'interprétation est conduite en réunissant les deux plans : le plan factoriel initial et le plan en étoile complémentaire. Pour cette interprétation les points au centre ne sont plus des points de contrôle. Ils sont pris en compte dans le calcul des coefficients.

■ Modélisation

Le modèle mathématique postulé est un modèle du second degré. Le même modèle est utilisé pour les deux réponses :

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

Tableau 9.8 – Coefficients du modèle (unités codées).

Coefficients	Rugosité × 1 000	Pics
a_0 (constante)	232,4	62,1
a_1	15,7	4,5
a_2	-65,5	3,7
a_{12}	-29,2	10
a_{11}	-39,2	-4,3
a_{22}	-21,8	19,6

Les coefficients du tableau 9.8 permettent d'écrire le modèle de la rugosité ($R^2 = 0,9993$) et celui du nombre de pics ($R^2 = 0,9937$) en unités codées :

$$\hat{y}_{\text{Rugosité}} = 232,4 + 15,7x_1 - 65,5x_2 - 29,2x_1x_2 - 39,2x_1^2 - 21,8x_2^2$$

$$\hat{y}_{\text{Pics}} = 62,1 + 4,5x_1 + 3,7x_2 + 10,0x_1x_2 - 4,3x_1^2 + 19,6x_2^2$$

■ Représentations graphiques des résultats

Les deux modèles établis peuvent servir à illustrer les résultats dans le domaine d'étude.

□ Analyse des résidus

On porte la valeur des résidus en fonction des réponses prévues (Figure 9.9). Les points semblent répartis au hasard et il n'apparaît pas de structures nettes comme sur la figure 5.10 du chapitre 5. Dans ces conditions, on considère qu'il n'y a plus d'information à extraire des données.

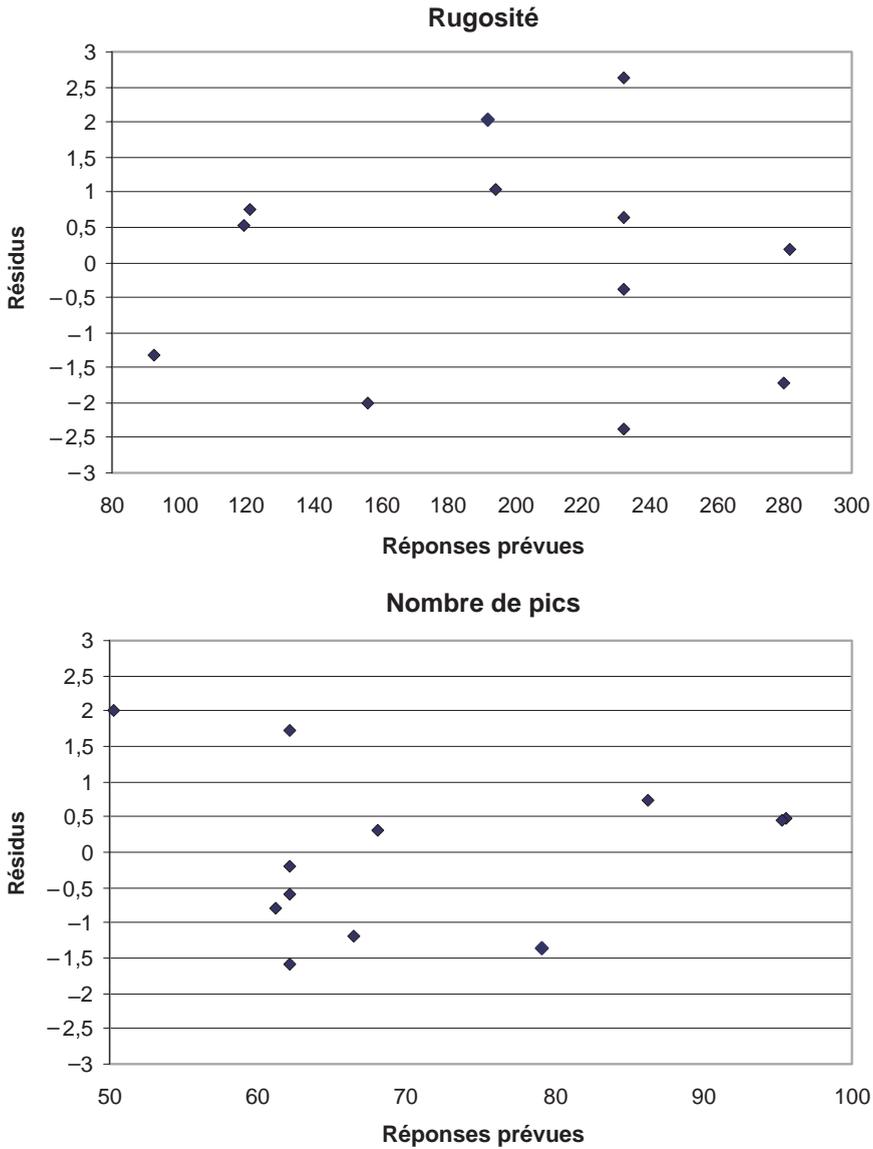


Figure 9.9 – Diagramme des résidus pour les modèles Rugosité et Pics.

□ Surfaces de réponse

Le modèle de la rugosité permet de tracer la surface de réponse correspondante (Figure 9.10). On constate que l'objectif d'une rugosité inférieure à 0,150 peut être atteint dans le domaine d'étude. Il suffit de choisir une vitesse d'avance et une vitesse de coupe qui, ensemble, donnent une réponse se situant sous la ligne de niveau 150. Le modèle des pics permet de tracer la surface de réponse (Figure 9.11). On constate que l'objectif d'un nombre de pics inférieur à 50 peut être atteint en limite du domaine d'étude.

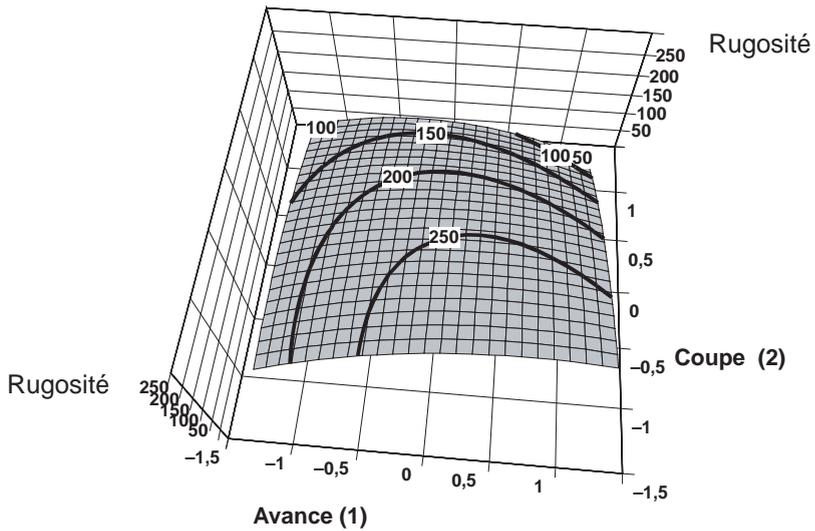


Figure 9.10 – Surface de réponse de la rugosité. Cette surface présente un maximum mais cela n'a pas d'intérêt ici puisque l'on cherche une rugosité inférieure à 0,150. (Les valeurs de rugosité ont été multipliées par 1 000.)

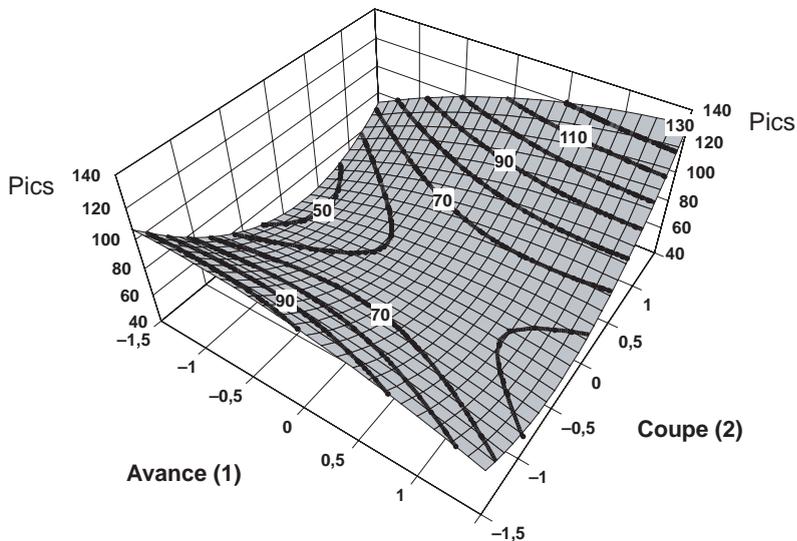


Figure 9.11 – Surface de réponse des pics. Un nombre de pics inférieur à 50 peut être atteint.

À l'examen des deux surfaces de réponse le contremaître s'aperçoit que son objectif peut être atteint. Il doit choisir un réglage de la vitesse de coupe et de la vitesse d'avance qui tienne compte des contraintes portant sur les deux réponses. Pour avoir une bonne précision, il trace les courbes d'isoréponses sur un même graphique (Figure 9.12).

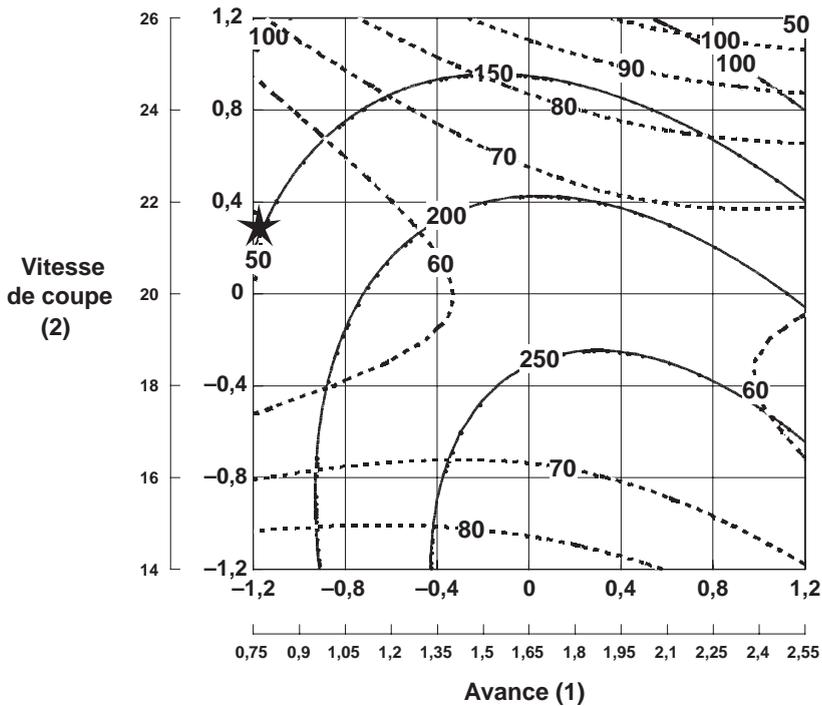


Figure 9.12 – Courbes d'isorréponses pour la rugosité (traits pleins) et pour le nombre de pics (traits pointillés).

Il constate qu'il existe une toute petite fenêtre correspondant (en unités centrées réduites) à :

$$\begin{aligned} \text{vitesse d'avance} &= -1,2 \\ \text{vitesse de coupe} &= 0,2 \end{aligned}$$

■ Confirmation des résultats

Comme ce réglage se trouve en extrémité de domaine, la précision de la prévision n'est pas bonne. Il ne peut pas encore donner de directives à ses ouvriers. Le contremaître décide de faire des essais de confirmation. Il calcule en unités normales, la vitesse de coupe et la vitesse d'avance correspondant aux valeurs centrées réduites :

$$\begin{aligned} \text{vitesse d'avance} &= 0,75 \text{ m/min} \\ \text{vitesse de coupe} &= 21 \text{ m/s} \end{aligned}$$

Il fait réaliser plusieurs pièces selon ces réglages et obtient des surfaces qui lui conviennent parfaitement. Il vérifie par des mesures de la rugosité et du nombre de pics que les prévisions des modèles sont satisfaites.

9.4.6 Conclusion de l'étude

L'état de surface des pièces rectifiées peut être amélioré si l'on choisit correctement les conditions d'usinage. Pour une profondeur de passe de 0,02 mm, la vitesse

d'avance de la meule doit être de 0,75 m/min et sa vitesse tangentielle doit être de 21 m/s. Ces réglages sont stricts car ils sont placés sur le flanc des surfaces de réponse et par suite instables. Ces réglages ne sont pas robustes et les ouvriers devront faire très attention et respecter les points de consigne. Une augmentation de la vitesse d'avance détériore l'état de surface. Une diminution de la vitesse d'avance améliorerait l'état de surface mais n'est pas économique. Une augmentation ou une diminution de la vitesse tangentielle de coupe détériore l'état de surface. Le contremaître décide donc de donner de nouvelles directives d'usinage qui devront être parfaitement respectées. Il appose la note suivante au tableau d'affichage :

Rectification : les conditions finales d'usinage devront être impérativement :

vitesse d'avance = 0,75 m/min

vitesse tangentielle de coupe = 21 m/s

profondeur de passe = 0,02 mm

9.5 Exemple 13 : un yoghourt doux (exemple de plan de Box-Behnken)

9.5.1 Préparation du plan d'expériences

■ Description de l'étude

Dans un laboratoire industriel un chercheur est chargé d'étudier l'influence de plusieurs facteurs susceptibles de diminuer l'acidité des yoghourts. L'acidité est produite par les ferments lactiques qui transforment le lactose en acide lactique. C'est la forte concentration de cet acide qui donne le goût particulier des yoghourts bulgares. L'objectif de l'étude est de diminuer le goût acide de ces laits fermentés. À cet effet, on prépare un lait stabilisé à partir d'un stabilisant naturel qui atténue les variations d'acidité du produit final, le yoghourt, malgré la présence des ferments lactiques.

La première étape du procédé (Figure 9.13) consiste à diluer le lait brut par de l'eau, opération nécessaire pour pouvoir traiter ensuite le lait brut.

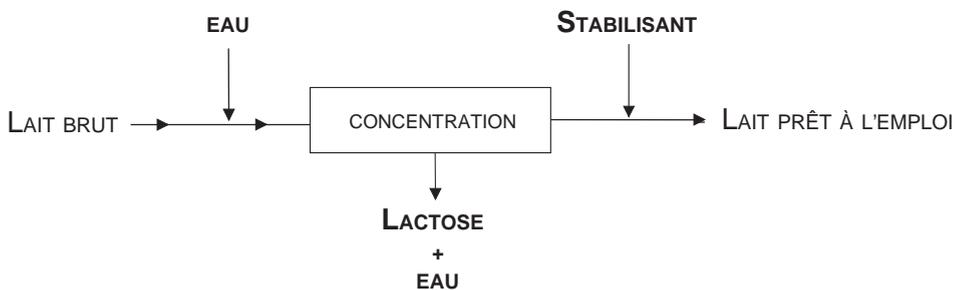


Figure 9.13 – Schéma de préparation du lait servant à fabriquer des yoghourts dont l'acidité est réduite.

La seconde étape du procédé est une concentration qui élimine une partie du lactose et de l'eau. Ayant moins de lactose à leur disposition, les ferments produiront moins d'acide lactique. Les deux premières étapes ont modifié le lait qui doit être stabilisé. C'est pourquoi le lait subit un nouveau traitement. On injecte un stabilisant. Cette injection ne modifie pas le volume du lait traité. On obtient, en fin de procédé, un lait stabilisé et prêt à l'emploi.

■ Facteurs

Les trois facteurs retenus par l'expérimentateur sont :

- Facteur 1 : taux de dilution. C'est le rapport volume d'eau ajoutée/volume de lait brut.
- Facteur 2 : pH lié à l'injection de stabilisant. On ajoute la quantité nécessaire de stabilisant pour obtenir un pH donné en fin d'injection. C'est le pH qui est contrôlé.
- Facteur 3 : taux de concentration du lait. C'est le rapport volume de lait brut/volume de lait stabilisé. Le volume de lait stabilisé est inférieur au volume de lait brut. Ce rapport est donc plus grand que l'unité.

■ Domaine d'étude

Les niveaux haut et bas de chaque facteur sont définis comme l'indique le tableau 9.9.

Tableau 9.9 – Domaine d'étude.

Facteur	Niveau -1	Niveau +1
Dilution (1)	0,5	2
pH (2)	6	5
Concentration (3)	1,5	2,5

■ Réponses

La réponse choisie par l'expérimentateur est l'« appauvrissement acide » qui mesure la perte en acide lactique. L'appauvrissement acide est d'autant meilleur que sa valeur est plus élevée. On cherchera donc des conditions de préparation du lait qui conduisent à une forte valeur de cette réponse. L'objectif est d'obtenir un lait stabilisé ayant un appauvrissement acide au moins égal à 48.

■ Choix du plan expérimental

L'expérimentateur s'attend à des variations de la réponse correspondant à un modèle du second degré. Après avoir examiné son problème en détail, il retient un plan de Box-Behnken qui permet d'établir le modèle désiré et ne nécessite qu'un nombre restreint d'essais.

9.5.2 Expérimentation

Les résultats des essais sont rassemblés dans le tableau 9.10. Les essais ont été reclassés dans l'ordre classique de présentation du plan de Box-Behnken. Les chiffres entre parenthèses indiquent l'ordre d'exécution des essais. Par exemple, l'essai n° 1 (5) est l'essai n° 1 du plan classique de Box-Behnken et il a été exécuté en cinquième.

Tableau 9.10 – Matrice d'expériences et résultats expérimentaux.

Essai n°	Dilution (1)	pH (2)	Concentration (3)	Réponses
1 (5)	–	–	0	51,3
2 (9)	+	–	0	42,6
3 (7)	–	+	0	42,2
4 (12)	+	+	0	50,4
5 (6)	–	0	–	40,7
6 (2)	–	0	+	41,5
7 (13)	+	0	–	41,3
8 (10)	+	0	+	40,8
9 (4)	0	–	+	35,2
10 (11)	0	+	–	35,3
11 (3)	0	–	–	39,5
12 (14)	0	+	+	39,8
13 (1)	0	0	0	50,8
14 (8)	0	0	0	50,1
15 (15)	0	0	0	49,4
Niveau –1	0,5	6	1,5	
Niveau 0	1,25	5,5	2	
Niveau +1	2	5	2,5	

9.5.3 Interprétation des résultats du plan

■ Modélisation

Le modèle mathématique postulé est un modèle du second degré :

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

L'interprétation commence par le calcul des coefficients du modèle (Tableau 9.11).

Tableau 9.11 – Coefficients du modèle (unités codées).

Coefficients	Appauvrissement	Coefficients	Appauvrissement
Constante a_0	50,10	a_{12}	4,22
		a_{13}	-0,32
a_1	-0,07	a_{23}	2,20
a_2	-0,11	a_{11}	0,07
a_3	0,06	a_{22}	-3,55
		a_{33}	-9,10

Ces coefficients permettent d'établir le modèle du second degré ($R^2 = 0,9968$) :

$$\hat{y} = 50,1 - 0,07x_1 - 0,11x_2 + 0,06x_3 + 4,22x_1x_2 - 0,32x_1x_3 + 2,2x_2x_3 + 0,07x_1^2 - 3,55x_2^2 - 9,1x_3^2$$

Les répétitions au centre permettent de calculer une estimation de l'erreur expérimentale (erreur pure). La somme des carrés de l'erreur expérimentale est 0,98 avec deux degrés de liberté (Tableau 9.12). On trouve un écart-type de 0,7.

La somme des carrés due au manque d'ajustement est de 0,3825 avec trois degrés de liberté. L'erreur d'ajustement est de 0,357, elle est donc plus faible que l'erreur pure. On ne pourra pas, statistiquement parlant, différencier ce modèle d'autres modèles mathématiques ayant, eux aussi, une erreur d'ajustement de l'ordre de grandeur de l'erreur expérimentale.

Tableau 9.12 – Comparaison du manque d'ajustement et de l'erreur expérimentale.

Origine des variations	Somme des carrés	Degrés de liberté	Carrés moyens
Manque d'ajustement	0,3825	3	0,1275
Erreur pure	0,9800	2	0,4900
Résidus	1,3625	5	0,2725

Analyse des résidus : la dispersion des résidus ne présente pas vraiment de tendance particulière (Figure 9.14). Rien de spécial n'apparaît sur ce graphique et les résidus semblent bien répartis au hasard.

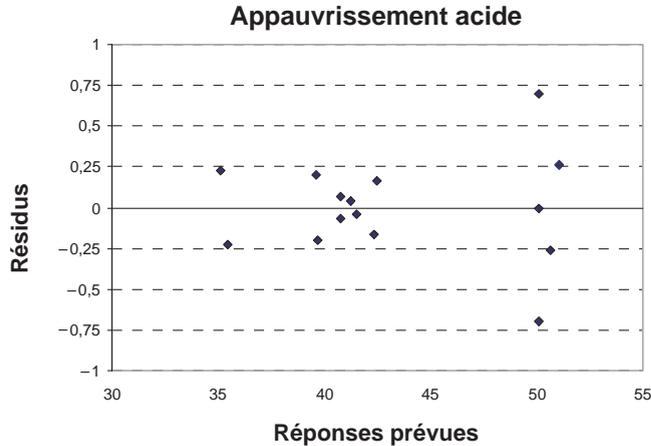


Figure 9.14 – Diagramme des résidus de la réponse « Appauvrissement acide ».

Le modèle que nous avons établi explique bien les résultats expérimentaux. Tous les tests d'évaluation le confirment. Mais, avant d'utiliser ce modèle, il faut le valider.

■ Interprétation des résultats

L'objectif est d'obtenir un appauvrissement acide supérieur à 48. En regardant le modèle, on voit immédiatement que cet objectif peut être atteint dans le domaine d'étude puisqu'au point central l'appauvrissement est de 50. On recherche la région du domaine d'étude où cette condition est respectée.

D'un point de vue économique, on a intérêt à choisir la dilution la plus faible possible, niveau -1 . C'est, en effet, pour ce niveau que l'on ajoute le moins d'eau pour une quantité donnée de lait brut. Par conséquent, les frais de fabrication seront moins élevés puisqu'il y aura moins d'eau à extraire. La dilution choisie est donc $0,5$.

Pour trouver le pH et la concentration, regardons la figure 9.15. Les courbes d'isoréponses indiquent qu'il y a une réponse maximale. Les coordonnées de ce maximum sont obtenues avec un logiciel de plans d'expériences :

$$\begin{cases} x_2 = -0,61 \\ x_3 = -0,05 \end{cases}$$

La réponse prédite en ce point est $51,60$.

Si l'on revient aux grandeurs naturelles, les conditions optimales sont les suivantes :

- Le taux de dilution sera réglé à $0,5$ (niveau -1). Cela signifie que l'on ajoutera, en début de traitement, un volume d'eau égal à la moitié du volume de lait brut.

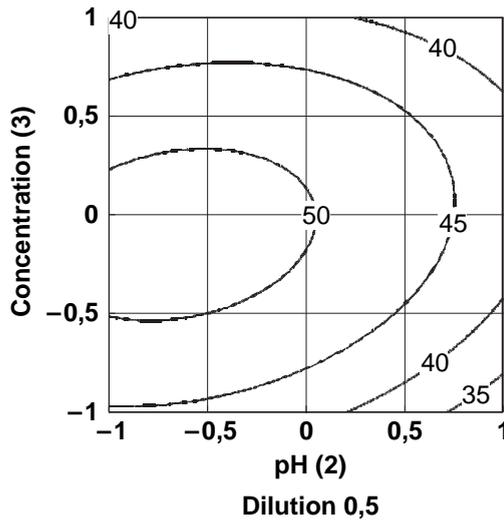


Figure 9.15 – Courbes d'isoreponses dans le plan pH-concentration pour une dilution de 0,5.

- Le pH sera fixé à 5,8 (niveau $-0,61$). On ajoutera la quantité nécessaire de stabilisant pour obtenir un pH de 5,8 en fin d'injection.
- Le taux de concentration du lait sera presque 2 (niveau $-0,05$). Le taux de concentration du lait est défini par le rapport (volume de lait brut/volume de lait stabilisé). Le volume de lait stabilisé sera égal à la moitié du volume de lait brut.

9.5.4 Conclusion de l'étude

Le lait stabilisé aura le meilleur appauvrissement acide possible si :

- on dilue le lait brut aux proportions suivantes : un volume d'eau pour deux volumes de lait brut ;
- on arrête l'injection de stabilisant à $\text{pH} = 5,8$.

Dans ces conditions, on obtiendra un volume de lait stabilisé pour deux volumes de lait brut et l'appauvrissement acide sera au moins de 50.

9.6 Exemple 14 : l'insecticide (exemple de plan de Doehlert)

9.6.1 Préparation du plan d'expériences

■ Description de l'étude

Un fabricant d'insecticide souhaite adapter son produit aux désirs de la clientèle. Il lui faut équilibrer son produit, c'est-à-dire mettre les justes concentrations des différents ingrédients pour obtenir les meilleurs résultats. Il étudie les trois produits de base qui constituent le produit commercial :

- l'insecticide lui-même : c'est le produit mortel qui tue les nuisibles ;

- le knock down : c'est une substance qui endort rapidement la vermine ;
- le synergiste : c'est un composé qui prépare et renforce l'action de l'insecticide.

Ces trois produits sont dilués dans un composé support qui les maintient en suspension et facilite leur vaporisation au moment de l'emploi. Un parfum est ajouté pour masquer l'odeur peu agréable des produits de base.

■ Facteurs et domaine d'étude

Le responsable retient, comme facteurs, les concentrations de chacun des produits de base :

- Facteur 1 : concentration en insecticide.
- Facteur 2 : concentration en knock down.
- Facteur 3 : concentration en synergiste.

Les niveaux haut et bas de chaque facteur sont définis comme l'indique le tableau 9.13.

Tableau 9.13 – Domaine d'étude.

Facteur	Niveau -1	Niveau 0	Niveau +1
Insecticide (1)	0,01 %	0,03 %	0,05 %
Knock down (2)	0,1 %	0,4 %	0,7 %
Synergiste (3)	0 %	1 %	2 %

■ Réponses

Les réponses choisies par l'expérimentateur sont le M24 et le KT50 :

- **M24** : c'est le pourcentage de nuisibles morts au bout de 24 heures. L'objectif est d'obtenir la plus grande valeur possible. On vise au moins 95 % et l'on cherchera plus si possible.
- **KT50** : c'est le temps, en minutes, au bout duquel 50 % de la population témoin est endormie. L'objectif est d'obtenir le plus faible KT50 possible. Un KT50 de 5 min ou moins serait excellent. Un KT50 compris entre 5 et 10 min serait un bon résultat.

■ Plans d'expériences

Les biologistes savent qu'il faut s'attendre à des surfaces de réponse du second degré. Un plan de Doehlert pour trois facteurs semble parfaitement convenir.

9.6.2 Expérimentation

Les trois facteurs « insecticide », « knock down » et « synergiste » sont étudiés. Les expérimentateurs prévoient trois points au centre. Les résultats des essais sont rassemblés dans le tableau 9.14.

Tableau 9.14 – Plan d'expériences et résultats.

Essai n°	Insecticide (1)	Knock down (2)	Synergiste (3)	KT50	M24
1	0	0	0	11	75
2	+1	0	0	15	57
3	+0,5	+0,866	0	11	56
4	-0,5	+0,866	0	6	65
5	-1	0	0	8	60
6	-0,5	-0,866	0	14	66
7	+0,5	-0,866	0	15	72
8	0	0	0	10	75
9	-0,5	+0,289	+0,816	2	91
10	0	-0,577	+0,816	8	99
11	+0,5	+0,289	+0,816	10	81
12	-0,5	-0,289	-0,816	11	72
13	0	+0,577	-0,816	8	81
14	+0,5	-0,289	-0,816	11	80
15	0	0	0	12	74
Niveau -1	0,01 %	0,1 %	0 %		
Niveau 0	0,03 %	0,4 %	1 %		
Niveau +1	0,05 %	0,7 %	2 %		

9.6.3 Interprétation des résultats du plan

■ Calcul des coefficients

On conduit les calculs pour chacune des deux réponses (Tableau 9.15 illustré par la Figure 9.16). On constate de très fortes courbures pour les coefficients a_{11} et a_{33} du M24 et pour le coefficient a_{33} du KT50.

Tableau 9.15 – Coefficients des modèles M24 et KT50 (unités codées).

Coefficients	M24	KT50
Constante a_0	74,7	11
a_1	-1,4	3,5
a_2	-4,8	-3,32
a_3	7,7	-2,04
a_{12}	-8,7	2,31
a_{13}	-7,9	4,08
a_{23}	-9,8	0,71
a_{11}	-16,2	0,50
a_{22}	-7,8	0,50
a_{33}	20	-4,25

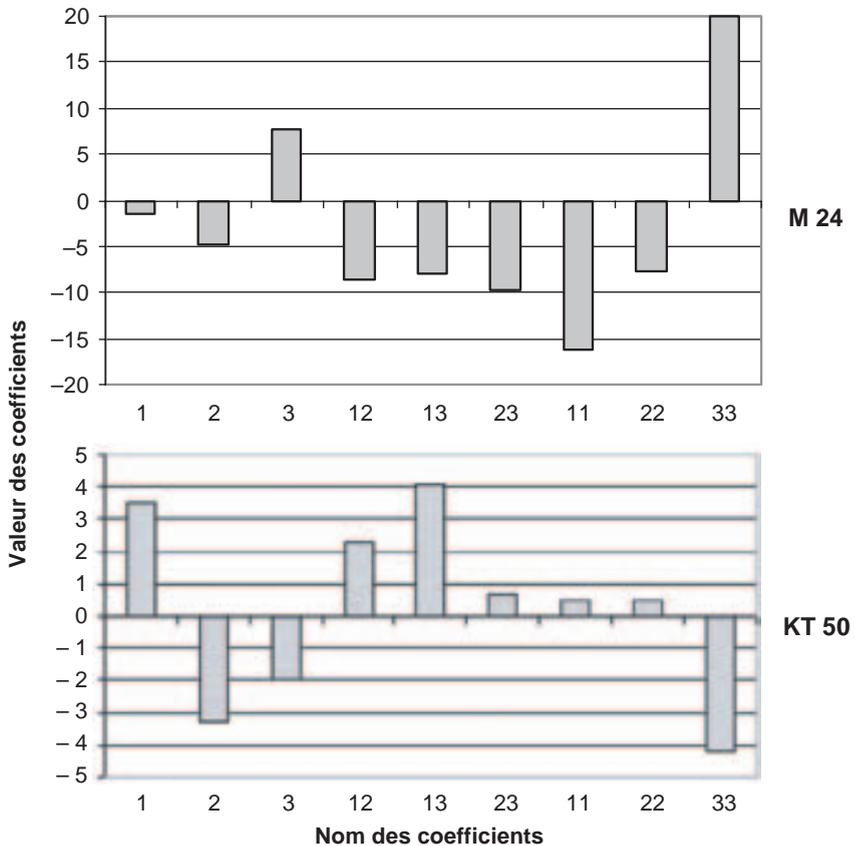


Figure 9.16 – Diagrammes à barres des coefficients des modèles du M24 et du KT50.

■ **Modélisation**

D'où les modèles (unités codées) du M24 ($R^2 = 0,9995$) et du KT50 ($R^2 = 0,983$) :

$$\hat{y}_{M24} = 74,7 - 1,4x_1 - 4,8x_2 + 7,7x_3 - 8,7x_1x_2 - 7,9x_1x_3 - 9,8x_2x_3 - 16,2x_1^2 - 7,8x_2^2 + 20x_3^2$$

$$\hat{y}_{KT50} = 11 + 3,50x_1 - 3,32x_2 - 2,04x_3 + 2,31x_1x_2 + 4,08x_1x_3 + 0,71x_2x_3 + 0,50x_1^2 + 0,50x_2^2 - 4,25x_3^2$$

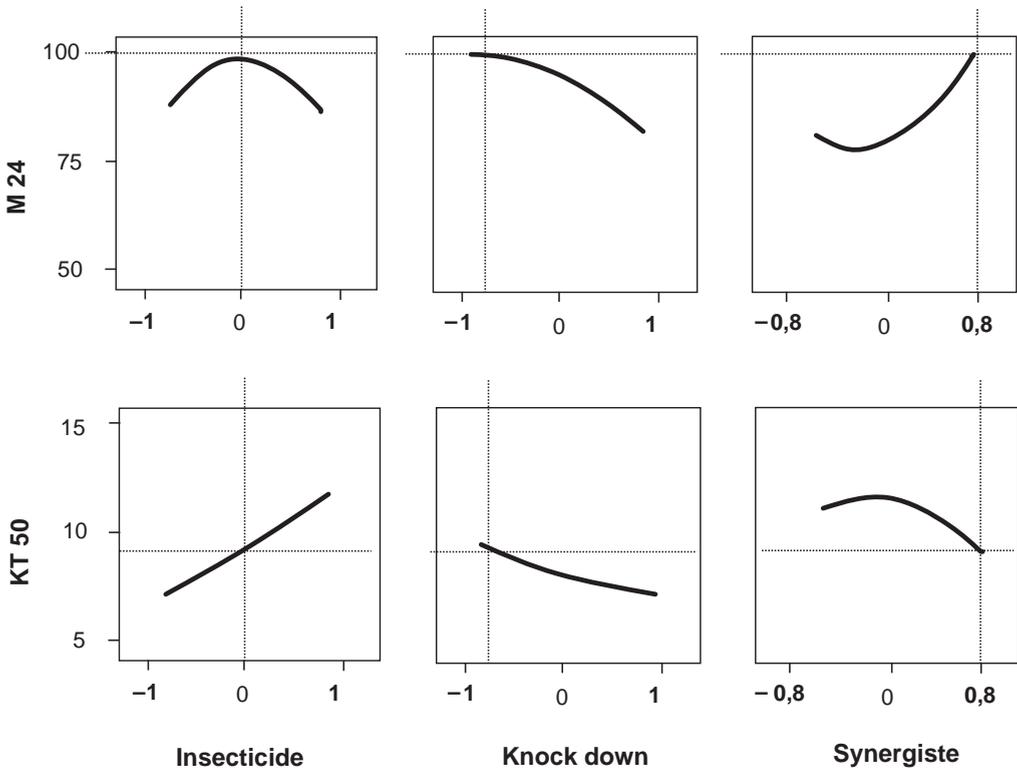


Figure 9.17 – Variation des réponses en fonction des facteurs.

On constate que le M24 est maximal pour une teneur en synergiste de +0,816, c'est-à-dire pour une concentration de 1,8 %. L'objectif étant de faire disparaître le plus possible de vermine, c'est cette concentration qui sera adoptée. Avant d'utiliser les modèles précédents, on effectue une analyse des résidus et l'on prévoit des expériences de contrôle pour vérifier la validité des conclusions.

■ Analyse des résidus

Rien d'anormal n'apparaît sur le graphique des résidus (Figure 9.18).

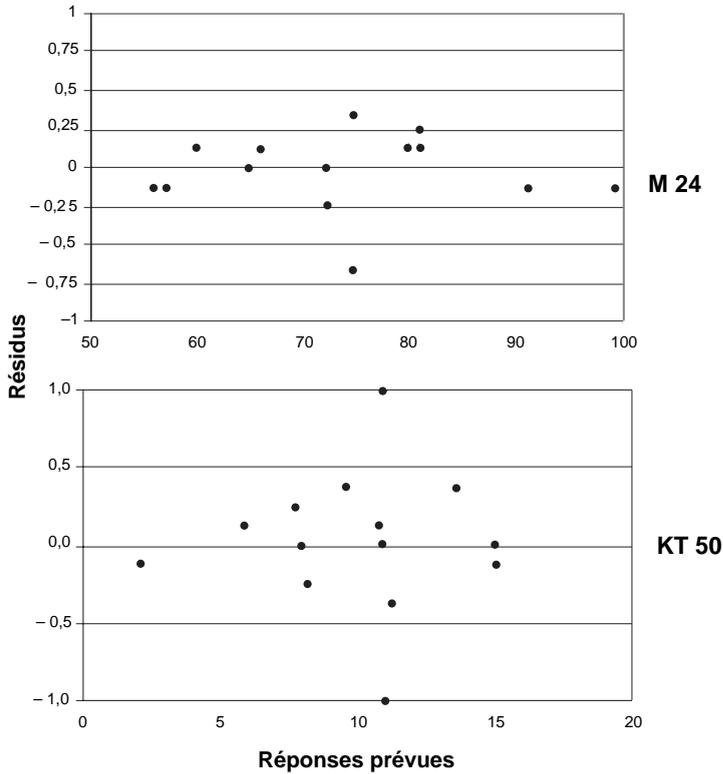


Figure 9.18 – Diagramme des résidus pour les modèles M24 et KT50.

■ Représentation graphique des résultats

Dans le modèle du M24, remplaçons x_3 par la valeur +0,816. On obtient ainsi le modèle du M24 pour le plan $x_3 = +0,816$:

$$\hat{y}_{M24} = 94,3 - 7,8x_1 - 12,8x_2 - 8,7x_1x_2 - 16,2x_1^2 - 7,8x_2^2$$

Ce modèle permet de tracer la surface de réponse (Figure 9.19). On constate que l'objectif peut être atteint. Les prévisions obtenues avec le modèle montrent que l'on peut dépasser une mortalité de 90 %. La mortalité la plus élevée est atteinte pour :

$$x_1 = -0,02 \text{ ou insecticide à } 0,03 \%$$

$$x_2 = -0,8 \text{ ou knock down à } 0,16 \%$$

La réponse est alors égale à 99,5 %.

Traçons la surface de réponse du KT50 (Figure 9.20) pour le même niveau de x_3 , soit +0,816 (1,8 % de synergiste). Le modèle est :

$$\hat{y}_{KT50} = 6,6 + 6,8x_1 - 2,7x_2 + 2,3x_1x_2$$

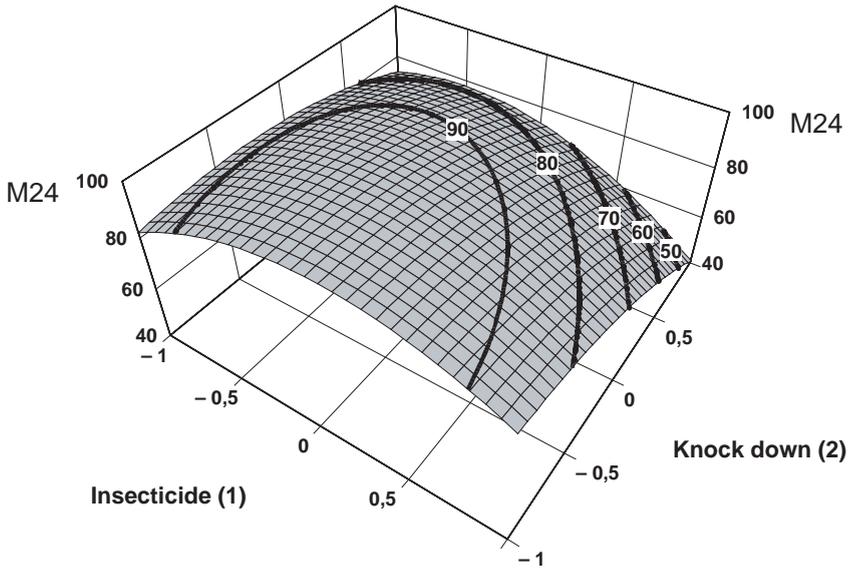


Figure 9.19 – Surface de réponse du M24 pour une concentration de 1,8 % en synergiste (niveau 0,816).

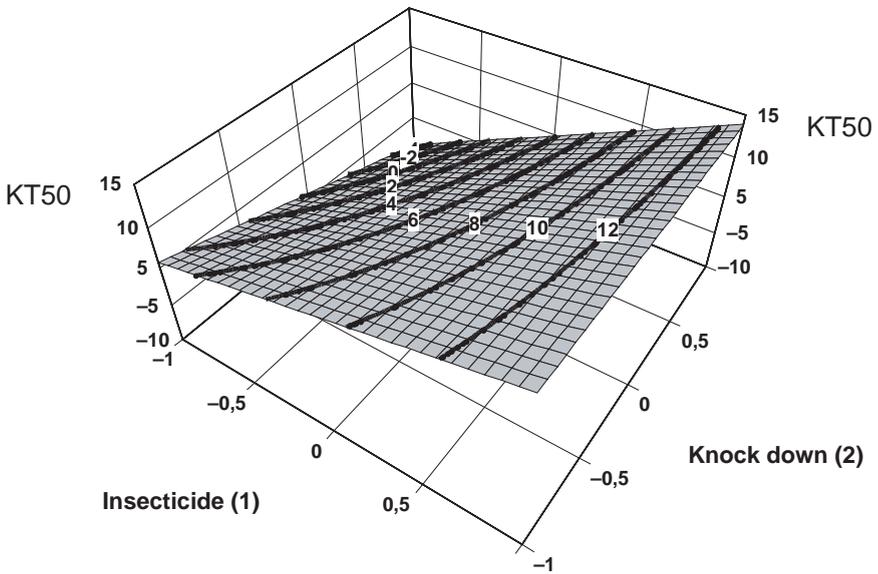


Figure 9.20 – Surface de réponse du KT50 pour une concentration de 1,8 % en synergiste.

On peut tracer sur un même graphique les courbes isoréponses pour le M24 et le KT50 (Figure 9.21). Le KT50 est légèrement inférieur à 10 minutes lorsque le M24 vaut 99,5 %.

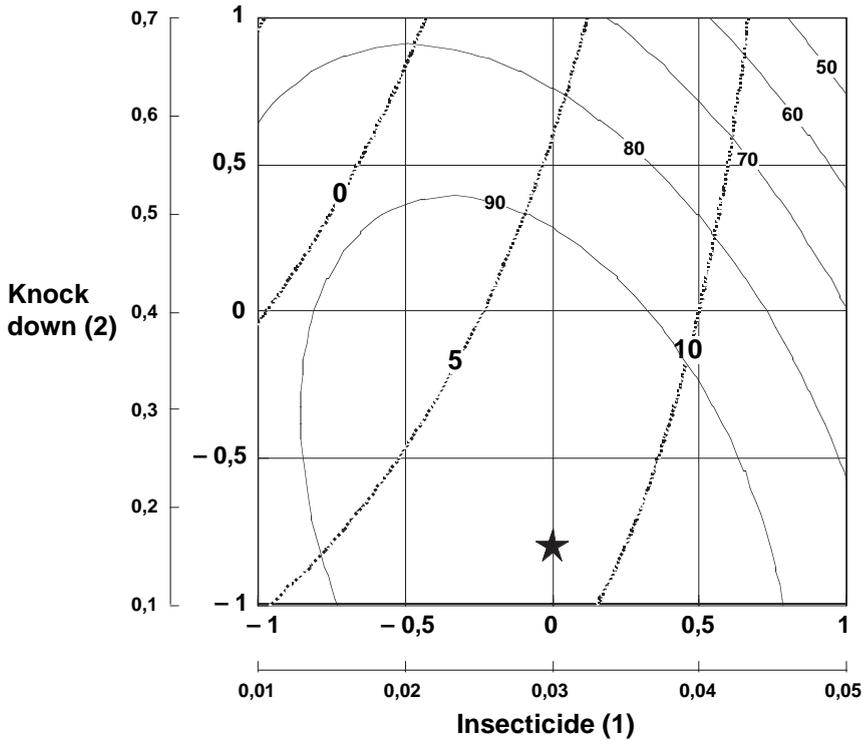


Figure 9.21 – Courbes d'isoréponses du KT50 (hyperboles en traits pointillés) et du M24 (ellipses en traits pleins) dans le plan $x_3 = 0,816$.

9.6.4 Conclusion de l'étude

Le modèle du M24 permet de préciser la région où l'on peut trouver une bonne mortalité. La composition du produit commercial, à ce stade de l'étude, est :

- Concentration en insecticide : 0,03 % (niveau 0).
- Concentration en knock down : 0,16 % (niveau -0,8).
- Concentration en synergiste : 1,80 % (niveau +0,8).

Avec cette composition on peut espérer un KT50 légèrement inférieur à 10 min et un M24 largement supérieur à 95 % (Figure 9.21). Les objectifs fixés au départ de l'étude peuvent être atteints. Des expériences de contrôle ont été lancées et ont confirmé les conclusions de l'étude.

10 • PLANS DE MÉLANGES

Dans les plans d'expériences classiques (plans factoriels, plans pour surfaces de réponse) les facteurs sont indépendants. Cela signifie que l'on peut choisir en toute liberté le niveau d'un facteur quels que soient les niveaux déjà attribués aux autres facteurs. Par exemple, si, pour un plan 2^4 , on a choisi les niveaux des trois premiers facteurs, on pourra encore choisir librement les niveaux du quatrième facteur. Cette liberté n'existe pas lorsque l'on étudie des mélanges car, en général, on étudie les réponses en fonction des proportions des constituants du mélange. Dans ce cas les facteurs d'étude sont les proportions des constituants du mélange. Comme la somme de ces proportions est toujours égale à 100 %, le pourcentage du dernier constituant est imposé par la somme des pourcentages des premiers composés du mélange. Dans cette situation les facteurs ne sont pas indépendants et cela entraîne des problèmes particuliers.

La méthodologie de mise en forme de l'étude et la conduite des essais sont tout à fait comparables à celles des plans d'expériences classiques.

Il faut bien se rendre compte qu'il y a un plan de mélanges lorsque la réponse étudiée dépend des proportions des constituants du mélange et non des quantités de mélange utilisées. Par exemple, la recette d'un cocktail est une indication de la proportion des différents ingrédients. La mise au point du cocktail en faisant varier les proportions des produits est la réalisation d'un plan de mélanges. En revanche, la consommation du cocktail en plus ou moins grande quantité (mais qui doit rester raisonnable car l'abus d'alcool nuit gravement à la santé) ne fait plus partie des plans de mélanges. Si la réponse dépend de la quantité du mélange, il s'agit alors de plans d'expériences classiques pour lesquels le choix des niveaux est libre.

Nous commencerons par examiner le problème de la non-indépendance des facteurs qui est à la base de la distinction entre les plans de mélanges et les plans d'expériences classiques. La non-indépendance des facteurs est exprimée par la *contrainte fondamentale des mélanges*.

10.1 Contrainte fondamentale des mélanges

Soit un mélange ayant n constituants. Le premier constituant représente un certain pourcentage du mélange, le second constituant un autre pourcentage du mélange, etc. Ainsi, chaque constituant participe pour une certaine part au mélange total. Mais l'ensemble des constituants du mélange forme un tout et la somme de leurs teneurs est égale à 100 %. Lorsque l'on a défini les proportions des $n - 1$ premiers constituants, la proportion du dernier ne peut plus être choisie, elle est déjà déterminée.

Si l'on note x_i la teneur en constituant i , la somme des teneurs de tous les constituants du mélange satisfait à la relation :

$$\sum_{i=1}^{i=n} x_i = 100 \%$$

La teneur de chaque constituant reste comprise entre 0 et 100 %. Lorsqu'on augmente la teneur de l'un des constituants, la teneur des autres est automatiquement diminuée pour que la somme de toutes les teneurs reste égale à 100 %.

Si, au lieu d'utiliser les pourcentages, on ramène la somme des teneurs des différents constituants à l'unité, on écrira :

$$\sum_{i=1}^{i=n} x_i = 1$$

Cette relation s'appelle la *contrainte fondamentale des mélanges*. C'est à cause de cette contrainte que les mélanges doivent être étudiés à part car les représentations géométriques des plans de mélanges sont différentes des représentations utilisées pour les plans d'expériences classiques et les modèles mathématiques sont eux aussi profondément modifiés.

10.2 Représentation géométrique des mélanges

10.2.1 Mélange à deux constituants

Soit x_1 la teneur du premier constituant et x_2 la teneur du second. Adoptons la représentation cartésienne : l'axe Ox_1 est orthogonal à l'axe Ox_2 . Les axes sont gradués en proportions variant de 0 à 1. Un mélange quelconque contenant x_a de A et x_b de B est représenté par un point situé à l'intersection des coordonnées x_a et x_b (Figure 10.1). Ce point qui représente un mélange est appelé soit *point de composition*, soit *point de mélange*, soit tout simplement *point*.

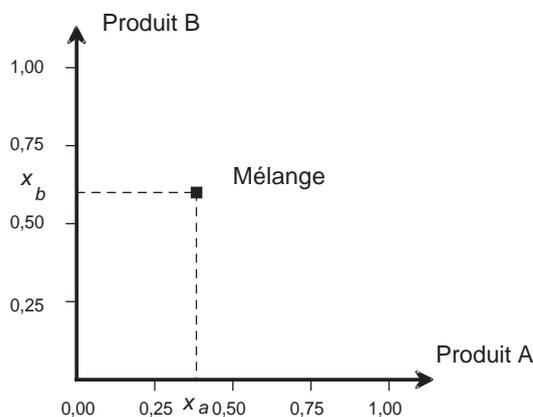


Figure 10.1 – Représentation d'un mélange dans un système d'axes cartésiens.

La contrainte des mélanges introduit une relation entre x_a et x_b :

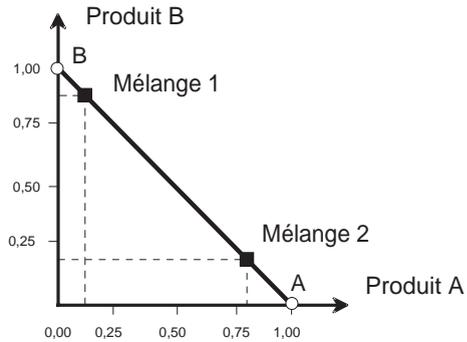
$$x_a + x_b = 1$$

que l'on peut écrire :

$$x_b = -x_a + 1$$

Cette dernière relation exprime que les points de coordonnées x_a et x_b sont sur une droite de pente -1 , coupant l'axe du produit A au point d'abscisse 1,00 (point A sur la figure 10.2) et l'axe du produit B au point d'ordonnée 1,00 (point B sur la figure 10.2). Les compositions x_a et x_b variant entre 0 et 1, il n'y a que le segment AB qui soit utile. Toutes les compositions possibles des mélanges des deux produits A et B sont représentées par les points de ce segment de droite.

Figure 10.2 – Les compositions des mélanges à deux constituants peuvent être représentées par les points du segment de droite AB.



On peut donc ne conserver que ce segment de droite et abandonner les axes Ox_1 et Ox_2 . Le produit pur A est représenté par l'une des extrémités du segment AB. Le produit B est représenté par l'autre extrémité. Ce segment porte une double graduation (Figure 10.3), celle des teneurs du premier produit et celle des teneurs du second produit. La lecture de la composition des mélanges sur ce segment demande un certain apprentissage. Elle se lit à la fois de droite à gauche et de gauche à droite.

10.2.2 Lecture du diagramme des mélanges binaires

Le segment de la figure 10.3 représente les mélanges des deux constituants A et B. Le produit A est à droite et le produit B est à gauche. La graduation inférieure est celle du produit A. Elle varie de 0 à gauche à 1 à droite. La graduation supérieure est celle du produit B. Elle varie de 0 à droite à 1 à gauche.

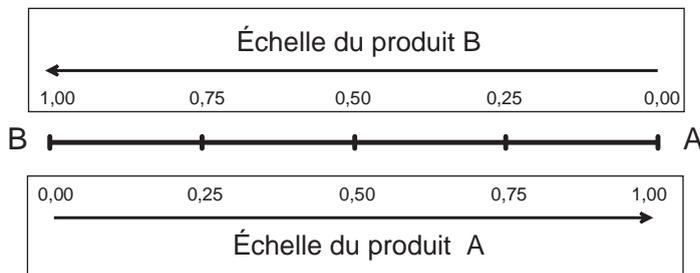


Figure 10.3 – Représentation des mélanges à deux constituants sur un segment de droite.

Les échelles se lisent, l'une dans un sens, l'autre dans l'autre sens. La somme des teneurs, en un point de l'échelle, est égale à l'unité.

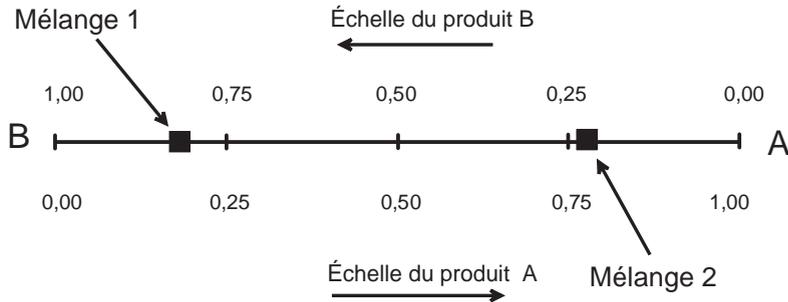


Figure 10.4 – Le mélange 1 contient 20 % de A et 80 % de B.
Le mélange 2 contient 77 % de A et 23 % de B.

Un premier mélange (mélange 1) est illustré par un point situé à 20 % de A et 80 % de B. Ce point est près de B et loin de A, ce qui est normal puisqu'il y a plus de B que de A dans le mélange (Figure 10.4). Un second mélange (mélange 2) est illustré par un point situé à 77 % de A et 23 % de B. Ce point est près de A puisque c'est le constituant le plus important. On peut supprimer l'une des deux échelles puisque la somme des teneurs en un point est toujours égale à 1. Connaissant l'une des teneurs, une simple soustraction donne la teneur complémentaire. C'est en général ce qui est fait.

10.2.3 Lecture du diagramme des mélanges ternaires

Un raisonnement semblable à celui que nous avons tenu pour deux constituants montre que l'on peut représenter les mélanges ternaires à l'aide d'un triangle équilatéral. Les produits purs sont aux sommets du triangle équilatéral. Les mélanges binaires sont représentés par les côtés du triangle. Par exemple le côté gauche AB du triangle (Figure 10.5) représente les compositions des mélanges des seuls produits A et B ; et ces mélanges ne contiennent pas de produit C.

Chaque côté du triangle équilatéral est gradué de 0 à 1. On n'a laissé que l'échelle d'un seul produit sur chaque côté. L'échelle de l'autre produit se déduit par complément à 1. Le côté AB est gradué pour le produit A : 1 est au point A et 0 au point B. C'est l'échelle du produit A. Le côté BC est gradué pour le produit B : 1 est au point B et 0 est au point C. Le côté CA est gradué pour le produit C : 1 est au point C et 0 est au point A.

Un point de la surface intérieure du triangle équilatéral représente un mélange ternaire. Les compositions de chaque produit se lisent sur les côtés du triangle. La teneur en produit A se lit sur le côté AB qui est l'échelle du produit A. On projette le point M sur AB parallèlement à BC (côté opposé à A). De même la teneur en B se lit sur l'échelle de B (côté BC) en projetant le point M sur BC parallèlement à AC (côté opposé à B). Enfin la teneur en C se lit sur l'échelle de C (côté AC) en projetant le point M sur à AC parallèlement à AB (côté opposé à C).

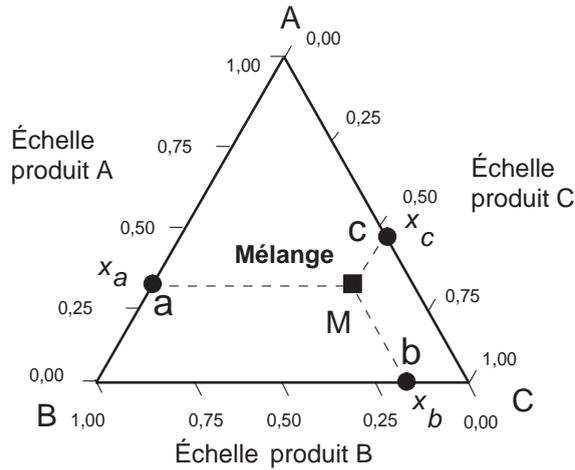


Figure 10.5 – Représentation des mélanges à trois constituants à l'aide d'un triangle équilatéral.

Les teneurs respectives des trois constituants dans le mélange sont données par les relations :

$$\begin{aligned} b_M &= B_a = \text{teneur en A dans le mélange M} \\ c_M &= C_b = \text{teneur en B dans le mélange M} \\ a_M &= A_c = \text{teneur en C dans le mélange M} \end{aligned}$$

Les propriétés géométriques du triangle équilatéral assurent le respect de la contrainte fondamentale des mélanges :

$$M_a + M_b + M_c = A_c + B_a + C_b = AB = AC = BC = 1$$

10.2.4 Mélanges à quatre constituants

Les quatre produits purs sont aux sommets d'un tétraèdre régulier (Figure 10.6). Les mélanges binaires sont représentés par les côtés du tétraèdre. Les mélanges ternaires sont représentés par les faces du tétraèdre qui sont des triangles équilatéraux. Les mélanges quaternaires sont représentés par les points du volume intérieur du tétraèdre. On obtient les compositions d'un mélange en projetant le point représentatif du mélange sur les faces et sur les côtés du tétraèdre.

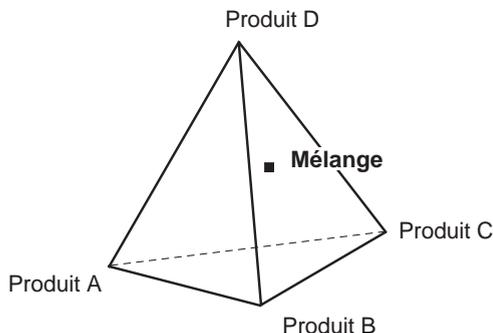


Figure 10.6 – Représentation des mélanges à quatre constituants par un tétraèdre régulier.

Afin de ne pas commettre d'erreur dans la lecture des compositions, on fera attention aux conventions qui ont été adoptées pour graduer les côtés du tétraèdre. Au-delà de quatre facteurs, la représentation géométrique n'est plus possible. On utilise alors uniquement la représentation en tableaux.

10.3 Plans de mélanges classiques

Les plans de mélanges classiques supposent que les produits purs ont la propriété d'usage recherchée et qu'il n'y a aucune contrainte qui pèse sur eux.

L'usage est de classer les principaux types de plans de mélanges suivant l'emplacement des points représentatifs des compositions. On distingue :

- les plans en réseaux,
- les plans de mélanges centrés,
- les plans de mélanges centrés augmentés.

10.3.1 Plans en réseaux

Les *plans en réseaux* (*Simplex lattice designs* dans la littérature anglo-saxonne) sont les premiers plans de mélanges qui ont été introduits. Ils ont été présentés par Scheffé dans les années 1958-1965. Les points de composition sont régulièrement repartis dans l'espace d'étude. Pour les mélanges à trois constituants, le plus simple des plans de mélanges en réseaux ne contient que les trois produits purs. Les points représentatifs de ces produits sont aux sommets du triangle équilatéral (Figure 10.7). Les coordonnées de ces points sont 0 et 1.

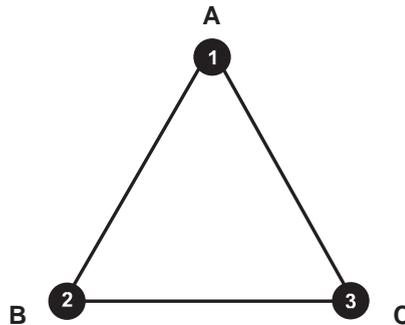


Figure 10.7 – Plan de mélanges ne comportant que les trois produits purs.

Le plan qui ne fait intervenir que les produits purs peut être enrichi par les points de composition moyenne, c'est-à-dire par les mélanges contenant 50 % de chacun des produits. Ces points de composition moyenne se situent au milieu des côtés du triangle. Leurs coordonnées sont 0 et 1/2.

Les points n° 4, 5 et 6 forment un réseau. La longueur unité des côtés du triangle a été divisée par 2 et les coordonnées de ces points sont égales à 1/2. Si l'on désire un réseau de points plus serré, il faut diviser la longueur des côtés par trois. Le pas de ce nouveau réseau est 1/3. On peut continuer à construire ainsi des réseaux de

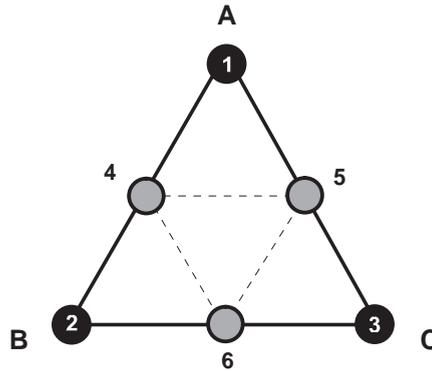


Figure 10.8 – Plan de mélanges en réseaux (*Simplex lattice designs*) comportant les trois produits purs et les mélanges moitié-moitié. Ce plan est noté {3,2}.

plus en plus serrés en divisant l'unité par m . Le pas du réseau est de $1/m$ et les coordonnées des points expérimentaux sont $0, 1/m, 2/m, 3/m$ jusqu'à $m/m = 1$.

Pour nommer ces plans on adopte la convention suivante : on utilise deux chiffres, le premier correspond au nombre de constituants du mélange et le second correspond au diviseur utilisé pour établir le pas du réseau. Ces deux chiffres sont séparés par une virgule et mis entre deux accolades. Ici, le plan de la figure 10.8 est un plan {3, 2}, soit 3 constituants et un pas de $1/2$. Le plan de la figure 10.9 est un plan {3, 3}, soit 3 constituants et un pas de $1/3$.

Cette notation permet de calculer très rapidement le nombre de compositions à préparer et à étudier connaissant le nombre de constituants du mélange et le nombre m . Pour un plan $\{q, m\}$, le nombre total de compositions différentes est égal à :

$$C_{q+m-1}^m = \frac{(q+m-1)!}{(m)!(q-1)!}$$

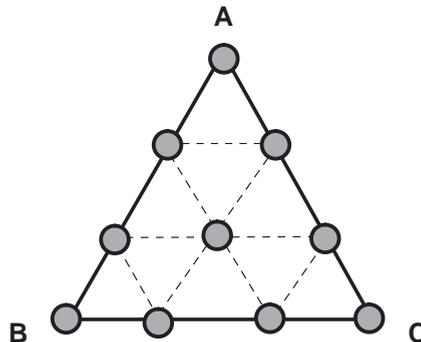


Figure 10.9 – Plan de mélanges en réseaux (*Simplex lattice designs*) comportant les trois produits purs et les mélanges $1/3$ - $2/3$ et $1/3$ - $1/3$ - $1/3$. Ce plan est noté {3, 3}.

10.3.2 Plans de mélanges centrés

Les *plans de mélanges centrés* (*Simplex-centroid designs* dans la littérature anglo-saxonne) se distinguent des plans en réseaux par la présence systématique d'un point central, c'est-à-dire d'une composition contenant autant de chacun des constituants du mélange (Figure 10.10). Le plan centré permettant d'étudier trois constituants comprend :

- les produits purs,
- les mélanges moitié-moitié de deux produits purs,
- le mélange contenant un tiers de chaque produit pur (point central).

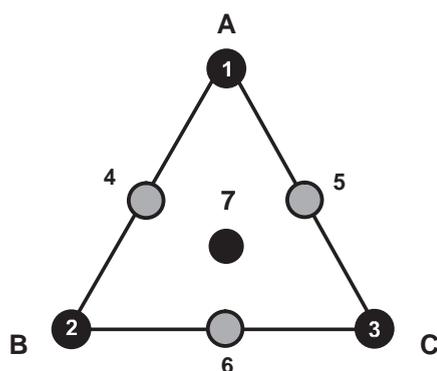


Figure 10.10 – Plan de mélanges centré (*Simplex-centroid designs*) comportant les trois produits purs, les mélanges moitié-moitié et le mélange équiportionnel des trois produits purs.

Le nombre N de compositions à étudier pour des plans de mélanges centrés à q constituants est donné par la formule :

$$N = 2^q - 1$$

Par exemple, dans le plan de mélanges centré à trois constituants, il y a 7 compositions différentes à étudier :

$$N = 2^3 - 1 = 8 - 1 = 7$$

10.3.3 Plans de mélanges centrés augmentés

Les *plans de mélanges centrés augmentés* (*Augmented Simplex-centroid designs* dans la littérature anglo-saxonne) sont des plans de mélanges centrés auxquels on ajoute les centres de gravité des simplex unitaires. Par exemple, le plan de mélanges centré pour trois composants comprend quatre simplex unitaires (Figure 10.11). Le centre de gravité de l'un des simplex unitaires est déjà occupé (point noir n° 7). Il reste donc à ajouter trois points (points blancs n° 8, 9 et 10) au centre des trois autres simplex unitaires.

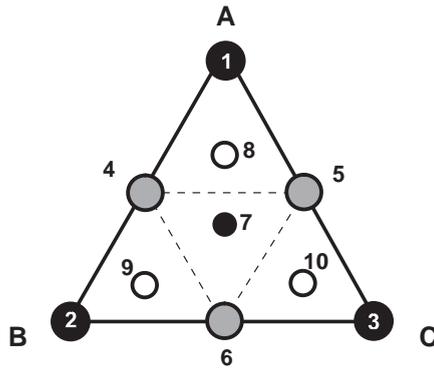


Figure 10.11 – Plan de mélanges centré augmenté (*Augmented Simplex-centroid designs*) comportant les trois produits purs, les mélanges moitié-moitié, le mélange équiportionnel des trois produits et les mélanges situés aux centres de gravité des simplex unitaires.

10.4 Modèles mathématiques des plans de mélanges

Le modèle mathématique appliqué aux plans de mélanges prend en compte la contrainte fondamentale des mélanges. Les polynômes utilisés présentent des particularités que nous allons indiquer.

10.4.1 Modèle du premier degré

On suppose que les variations de la réponse sont proportionnelles aux compositions du mélange. On pourrait adopter, en premier lieu, le même modèle que celui des plans factoriels classiques dont les facteurs sont indépendants. Par exemple, pour un mélange à trois constituants on pourrait écrire en un point donné :

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 \quad \{10.1\}$$

Mais les compositions x_i ne sont pas indépendantes. Il faut tenir compte de la contrainte fondamentale des mélanges :

$$x_1 + x_2 + x_3 = 1$$

La relation {10.1} peut alors s'écrire :

$$y = a_0(x_1 + x_2 + x_3) + a_1x_1 + a_2x_2 + a_3x_3$$

ou en regroupant les coefficients :

$$y = (a_0 + a_1)x_1 + (a_0 + a_2)x_2 + (a_0 + a_3)x_3$$

Ce modèle ne comporte plus de terme constant. Si l'on change de notation et que l'on pose :

$$b_1 = a_1 + a_0 \quad b_2 = a_2 + a_0 \quad b_3 = a_3 + a_0$$

le modèle prend alors la forme suivante :

$$y = b_1x_1 + b_2x_2 + b_3x_3 \quad \{10.2\}$$

Il n'y a pas de terme constant. Cette disparition est la conséquence directe de la contrainte fondamentale des mélanges.

10.4.2 Modèle du second degré

Le modèle mathématique du second degré comprend les termes du premier degré, les termes rectangles et les termes carrés. Comme précédemment, nous allons partir du modèle mathématique des plans factoriels et nous allons examiner les conséquences de la contrainte fondamentale des mélanges. Pour un mélange à deux constituants, on pourrait écrire sachant qu'il n'y a pas de constante :

$$y = a_1x_1 + a_2x_2 + a_{12}x_1x_2 + a_{11}x_1^2 + a_{22}x_2^2 \quad \{10.3\}$$

Prenons en compte la contrainte fondamentale des mélanges :

$$x_1 + x_2 = 1$$

que l'on peut écrire :

$$x_1 = 1 - x_2$$

Multiplions chaque membre par x_1 :

$$x_1^2 = x_1(1 - x_2)$$

$$x_1^2 = x_1 - x_1x_2$$

On constate que le terme carré est converti en un terme du premier degré et en un terme rectangle. Les termes carrés disparaissent donc de la relation initiale. On obtient un modèle comprenant uniquement des termes du premier degré et des termes rectangles, soit en adaptant les notations :

$$y = b_1x_1 + b_2x_2 + b_{12}x_1x_2 \quad \{10.4\}$$

Pour un mélange à trois constituants, on a :

$$y = b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3$$

10.4.3 Modèle du troisième degré

Le modèle mathématique du troisième degré, particulier aux plans de mélanges, peut s'obtenir à partir d'un polynôme du troisième degré sur lequel on applique la contrainte fondamentale des mélanges. Le modèle, appelé *modèle cubique complet*, est le suivant :

$$y = b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 \\ + b'_{12}x_1x_2(x_1 - x_2) + b'_{13}x_1x_3(x_1 - x_3) + b'_{23}x_2x_3(x_2 - x_3) + b'_{123}x_1x_2x_3$$

Il est plus courant d'utiliser le modèle simplifié qui ne comprend que les termes du premier degré, les termes rectangles et le produit des trois constituants. Le modèle simplifié, appelé *modèle cubique restreint*, correspondant à un mélange de trois constituants est le suivant :

$$y = b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3 \quad \{10.5\}$$

10.5 Exemple 15 : les trois polymères

10.5.1 Préparation du plan d'expériences

■ Description de l'étude

Un expérimentateur cherche la relation qui existe entre l'allongement à la rupture des fils fabriqués avec trois polymères (polyéthylène, polystyrène et polypropylène) et la composition de leurs mélanges. Le modèle du premier degré étant insuffisant, pour expliquer les résultats expérimentaux, l'expérimentateur va tester le modèle du second degré.

■ Réponse

La réponse est l'allongement à la rupture des fils préparés avec les différents mélanges de polymères.

■ Facteurs et domaine

Les facteurs sont les proportions de chacun des trois polymères polyéthylène, polystyrène et polypropylène.

Les proportions peuvent varier de 0 à 100 % car la propriété d'usage, l'allongement à la rupture, est valable aussi bien pour les produits purs que pour les mélanges. Le domaine d'étude est représenté par tout le triangle équilatéral.

■ Plan

Le plan choisi est un plan de mélanges centré augmenté (*Augmented Simplex-centroid designs*) qui comporte les mélanges indiqués dans le tableau 10.1. L'expérimentateur prévoit d'utiliser les points n° 8, 9 et 10 comme points de contrôle (Figure 10.12). Ces points ne seront pas utilisés pour établir le modèle mathématique, ils serviront à vérifier la valeur prédictive du modèle.

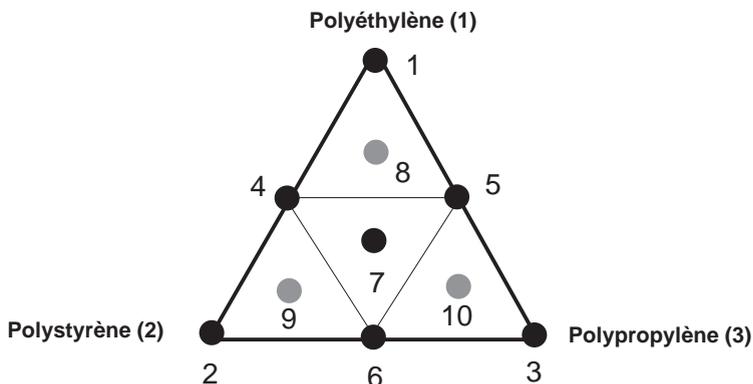


Figure 10.12 – Position des points expérimentaux (noirs) et des points de contrôle (gris) du plan utilisé pour étudier l'allongement à la rupture des mélanges de polymères.

10.5.2 Expérimentation

Les résultats des mesures figurent dans le tableau 10.1.

Tableau 10.1 – Plan d'expériences et points de contrôle.

Mélange n°	Polyéthylène (1)	Polystyrène (2)	Polypropylène (3)	Réponses
1	1	0	0	32
2	0	1	0	25
3	0	0	1	42
4	1/2	1/2	0	38
5	1/2	0	1/2	39
6	0	1/2	1/2	30,5
7	1/3	1/3	1/3	37
8	2/3	1/6	1/6	37
9	1/6	2/3	1/6	32
10	1/6	1/6	2/3	38

10.5.3 Interprétation des résultats

■ Calcul des coefficients et modélisation

Un plan centré augmenté pour trois constituants permet de calculer les sept coefficients du modèle cubique restreint. Les trois points de contrôle ne sont pas intégrés au calcul de ces coefficients. Le tableau 10.2 donne la valeur des sept coefficients.

Tableau 10.2 – Coefficients du modèle cubique restreint (unités codées).

Coefficient	Valeur
b_1	32
b_2	25
b_3	42
b_{12}	38
b_{13}	8
b_{23}	-12
b_{123}	6

On obtient ainsi les valeurs numériques des coefficients du modèle mathématique. Ces coefficients permettent d'écrire un modèle avec lequel on peut calculer les réponses dans l'ensemble du domaine d'étude :

$$\hat{y} = 32x_1 + 25x_2 + 42x_3 + 38x_1x_2 + 8x_1x_3 - 12x_2x_3 + 6x_1x_2x_3$$

■ Examen de la validité du modèle cubique restreint

L'expérimentateur a prévu trois mélanges de contrôle : les points n° 8, 9 et 10 (Tableau 10.3). Calculons avec le modèle la valeur des réponses prévues en ces points. Puis comparons ces valeurs aux réponses mesurées.

Tableau 10.3 – Comparaison des réponses aux points de contrôle.

Mélange n°	Réponses mesurées	Réponses calculées
8	37	37,38
9	32	32,22
10	38	38,22

La comparaison de la réponse mesurée et de la réponse calculée montre que l'on peut considérer ces valeurs comme égales. On en conclut que le modèle cubique restreint représente correctement l'allongement à la rupture des fils fabriqués avec des mélanges des trois polymères.

■ Illustration graphique des résultats

Maintenant que le modèle est validé, nous pouvons l'utiliser pour tracer les courbes isoréponses. On a ainsi une idée de l'évolution de la grandeur d'intérêt en fonction des proportions de chacun des constituants (Figure 10.13).

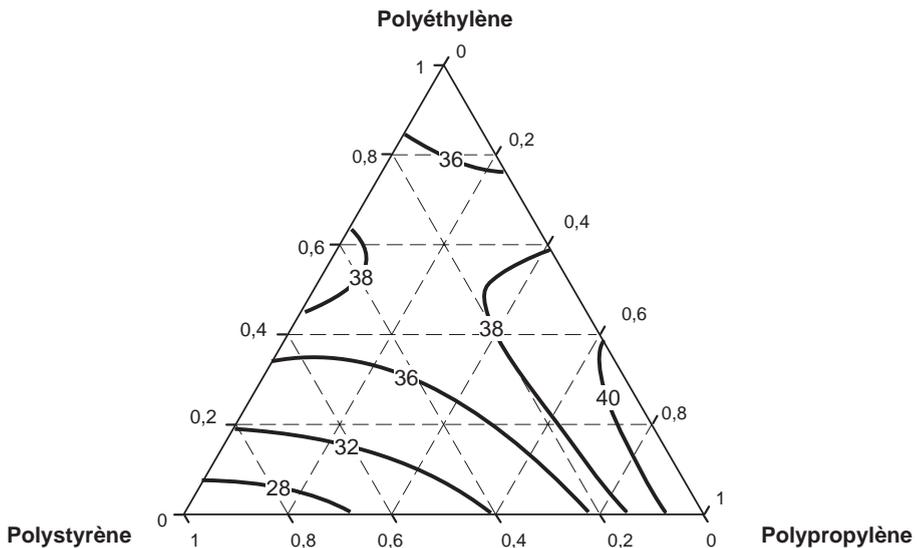


Figure 10.13 – Courbes isoréponses (allongement à la rupture) tracées à l'aide du modèle du second degré.

10.5.4 Conclusion de l'étude

L'objectif de connaissance est atteint. On peut calculer l'allongement à la rupture des mélanges de polyéthylène-polystyrène-polypropylène pour toutes les compositions du domaine d'étude. Si l'on vise une forte valeur, il faudra choisir des mélanges à fortes teneurs en polypropylène et à très faible teneur en polystyrène. Si on désire un allongement peu sensible aux variations de composition, on se placera au point stationnaire : environ 20 % de polypropylène, 30 % de polystyrène et 50 % de polyéthylène. Si on cherche un faible allongement à la rupture, on choisira des mélanges riches en polystyrène.

11 • NOTION DE PLAN OPTIMAL

Dans les chapitres précédents, nous avons étudié les plans d'expériences classiques. Le nombre d'essais dépendait du plan choisi. Après quelques plans d'expériences, l'expérimentateur se demande s'il ne pourrait pas choisir le nombre d'essais plutôt que le plan lui-même. Et quand il considère l'aspect économique, la question devient : « Quel est le nombre minimal d'essais dont j'ai besoin pour traiter mon problème ? » Le présent chapitre est une introduction aux plans d'expériences optimaux qui apportent une réponse à cette question posée. Ces plans portent aussi le nom de *custom designs* dont la traduction pourrait être *plans sur mesure* pour bien montrer qu'ils sont conçus pour s'adapter à l'étude et aux exigences de l'expérimentateur.

11.1 Exemple d'Hotelling

Pour illustrer les problèmes soulevés par le choix d'une stratégie expérimentale nous allons prendre l'exemple des pesées de Hotelling. En effet, la première personne qui s'est posé la question de l'optimisation des pesées est Hotelling qui a publié en 1944 une étude sur ce sujet (voir bibliographie). On peut être surpris que ce problème d'optimisation n'ait été abordé qu'en 1944 dès lors que l'on pèse depuis des milliers d'années.

Supposons qu'un expérimentateur possède un trébuchet analogue à celui représenté sur la figure 11.1 et qu'il a deux objets A et B dont il veut déterminer les masses m_a et m_b .



Figure 11.1 – Balance à deux plateaux.

Sur un plateau, il met l'objet A et, sur l'autre, il ajoute des poids jusqu'à ce que l'aiguille revienne à sa position d'équilibre initiale (Figure 11.2). Soit p_1 le total des poids à ce moment. Comme l'expérimentateur sait qu'il commet une erreur σ sur p_1 il écrit :

$$m_a = p_1 + \sigma$$

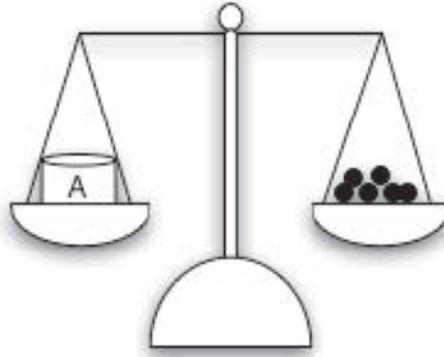


Figure 11.2 – Représentation de la pesée de l'objet A.

Pour le deuxième objet, il opère de même et trouve qu'il faut mettre un poids p_2 pour retrouver l'équilibre de la balance :

$$m_b = p_2 + \sigma$$

Si p_1 et p_2 valent respectivement 10 g et 25 g et si σ vaut 0,1 g, l'expérimentateur peut écrire :

$$m_a = 10 \pm 0,1 \text{ g}$$

$$m_b = 25 \pm 0,1 \text{ g}$$

L'expérimentateur a fait deux mesures et il a obtenu deux résultats avec une erreur de $\pm \sigma$. C'est ainsi que les gens pratiquent dans tous les métiers du monde. Pourtant, Hotelling a fait remarquer que, pour le même effort, on pouvait obtenir une meilleure précision, à condition que les deux objets participent à chaque pesée. Il propose donc dans une première expérience de mettre A et B sur un même plateau, c'est-à-dire de mesurer la somme p_1 des masses. Puis, dans une seconde expérience de mettre A et B chacun sur un plateau et d'obtenir ainsi la différence p_2 des masses de A et de B. Il peut ainsi écrire un système d'équations dont il déduit les masses m_a et m_b par un calcul simple :

$$m_a + m_b = p_1$$

$$m_a - m_b = p_2$$

d'où :

$$m_a = \frac{1}{2}(p_1 + p_2)$$

$$m_b = \frac{1}{2}(p_1 - p_2)$$

Si on applique le théorème des variances (voir chapitre 5), on obtient :

$$\begin{aligned} V(m_a) &= \frac{1}{4}[V(p_1) + V(p_2)] \\ &= \frac{1}{4}[\sigma^2 + \sigma^2] \\ &= \frac{1}{2}\sigma^2 \end{aligned}$$

L'erreur (l'écart-type) sur les résultats de mesures n'est plus que $\sigma/\sqrt{2}$ au lieu de σ , c'est-à-dire une réduction de 30 % de l'erreur de mesure. On écrit :

$$\begin{aligned} m_a &= 10 \pm 0,07 \text{ g} \\ m_b &= 25 \pm 0,07 \text{ g} \end{aligned}$$

Avec la méthode classique, le même résultat aurait été atteint en pesant deux fois A et en pesant deux fois B, soit 4 mesures au lieu de 2. Ici commence à poindre l'un des intérêts de faire intervenir toutes les variables, ou tous les facteurs, à chaque expérience : pour un même nombre d'expériences on améliore la précision.

11.2 Pesées et plans d'expériences

Reprenons l'exemple d'Hotelling et traitons-le avec le formalisme des plans d'expériences introduit aux chapitres 1 et 2.

Les facteurs seront les objets A et B et les réponses les poids nécessaires pour ramener l'équilibre à chaque pesée. On adopte une convention de signe : si l'objet est déposé sur le plateau de gauche, il est compté positivement, s'il est déposé sur le plateau de droite, il est compté négativement (Figure 11.3).

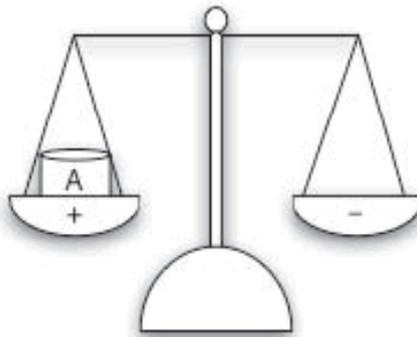


Figure 11.3 – Convention de signes adoptée pour les pesées.

Avec cette convention la masse de l'objet A est $-p_a$ si A est sur le plateau de droite et $+p_a$ si A est sur le plateau de gauche. On peut simplifier la notation en introduisant les variables codées, -1 correspond à $-p_a$, $+1$ à $+p_a$ et 0 à l'absence de A :

$$\begin{array}{ccc} +p_a & \text{—} & 0 & \text{—} & -p_a \\ +1 & \text{—} & 0 & \text{—} & -1 \end{array}$$

Les mêmes conventions sont adoptées pour l'objet B. L'effet de B sur les plateaux s'exprime par :

$$\begin{array}{ccc} +p_b & \text{—} & 0 & \text{—} & -p_b \\ +1 & \text{—} & 0 & \text{—} & -1 \end{array}$$

Les deux manières de peser, la méthode classique et la méthode d'Hotelling peuvent alors être présentées selon une matrice d'expériences. Les calculs seront conduits comme pour un plan d'expériences et les effets trouvés seront les poids de chacun des objets : l'effet d'un objet étant, dans ce cas, son poids.

11.2.1 Méthode classique

- Essai n° 1 : A est sur le plateau de gauche, B n'est pas sur la balance. Réponse : 10 g
- Essai n° 2 : B est sur le plateau de gauche, A n'est pas sur la balance. Réponse : 25 g

Tableau 11.1 – Matrice d'expériences : méthode classique pour peser deux objets.

Essai n°	A Facteur 1	B Facteur 2	Masse (g) Réponses
1	+1	0	10
2	0	+1	25
Effet	10 g	25 g	

Comme il n'y a aucune interaction entre A et B, la matrice d'expériences (Tableau 11.1) est identique à la matrice du modèle. La matrice \mathbf{X} de la méthode classique de pesée est donc :

$$\mathbf{X} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \{11.1\}$$

Il ne faut pas s'étonner de voir les effets égaux aux réponses : la matrice des effets est une matrice unité, elle n'apporte aucune transformation.

11.2.2 Méthode d'Hotelling

- Essai n° 1 : A et B sont sur le plateau de gauche. Réponse : 35 g
- Essai n° 2 : B est sur le plateau de gauche, A sur le plateau de droite. Réponse : 15 g

La matrice \mathbf{X} de la méthode d'Hotelling (Tableau 11.2) est :

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

Tableau 11.2 – Matrice d'expériences : méthode d'Hotelling pour peser deux objets.

Essai n°	A Facteur 1	B Facteur 2	Masse (g) Réponses
1	+1	+1	35 g
2	-1	+1	15 g
Effet	10 g	25 g	

La méthode d'Hotelling est caractérisée par cette matrice où ne figure aucun zéro, ce qui signifie que tous les objets ont participé à tous les essais.

On retrouve, bien sûr, les mêmes effets (ou les mêmes poids) qu'avec la méthode classique, avec cependant, et la différence est de taille, une meilleure précision.

11.3 Optimalité

Nous venons de voir que la méthode d'Hotelling donne des résultats plus précis que la méthode classique. La question qui se pose naturellement est : « Peut-on faire mieux ? Peut-on obtenir avec deux essais des effets plus précis ? La méthode d'Hotelling est-elle optimale ? »

Nous adopterons comme définition : « *un plan optimal est un plan qui fournit les estimations des effets avec la meilleure précision possible* ». Les mathématiciens ont démontré que ce but était atteint lorsque le déterminant de la matrice $\mathbf{X}'\mathbf{X}$ atteignait sa valeur maximale. Le calcul du déterminant d'une matrice est fort long sauf pour les plus petites matrices. C'est la raison pour laquelle on fait appel aux logiciels et aux ordinateurs pour générer les matrices des plans optimaux. Pour illustrer la méthode, nous utiliserons un exemple très simple.

11.3.1 Critère du déterminant maximum (D-optimal)

Indiquons d'abord, d'une manière mathématique, le critère d'optimalité qui fait intervenir le déterminant d'une matrice. On notera $|\mathbf{A}|$ le déterminant de la matrice \mathbf{A} . On dira qu'un plan d'expériences est *D-optimal* si la position des points d'expériences maximise la valeur du déterminant $|\mathbf{X}'\mathbf{X}|$ de la matrice $\mathbf{X}'\mathbf{X}$. Le D de D-optimal vient de *déterminant* et on parle de *D-optimalité*.

Les déterminants ne peuvent être calculés que pour les matrices carrées. Mais comme les matrices $\mathbf{X}'\mathbf{X}$ sont toujours carrées, on peut toujours calculer leurs déterminants.

Appliquons cette notion de D-optimalité à l'exemple des pesées, avec la méthode classique et la méthode d'Hotelling :

– pour la méthode classique :

$$\mathbf{X} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{et} \quad |\mathbf{X}'\mathbf{X}| = 1$$

– pour la méthode d’Hotelling :

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \quad \text{et} \quad |\mathbf{X}'\mathbf{X}| = 4$$

La stratégie d’Hotelling est meilleure que la stratégie classique car le déterminant de $|\mathbf{X}'\mathbf{X}|$ est plus élevé. En fait, on démontre que l’on ne peut pas trouver une meilleure stratégie que celle d’Hotelling.

11.3.2 Calcul des plans optimaux

La stratégie d’Hotelling peut être étendue à la pesée de quatre objets A, B, C et D. Comment faut-il organiser les pesés pour obtenir la meilleure précision ? Étudions les différentes stratégies possibles.

Voyons d’abord la méthode classique sachant qu’elle n’est pas optimale. On peut représenter les quatre pesées par la matrice :

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Le déterminant de la matrice $\mathbf{X}'\mathbf{X}$ est égal à :

$$|\mathbf{X}'\mathbf{X}| = 1$$

On peut mettre en œuvre une approche analogue à celle qui a été utilisée pour deux objets. On pèse d’abord A et B puis C et D selon les indications suivantes :

- Essai n° 1 : A et B sont sur le plateau de gauche.
- Essai n° 2 : B est sur le plateau de gauche, A sur le plateau de droite.
- Essai n° 3 : C et D sont sur le plateau de gauche.
- Essai n° 4 : D est sur le plateau de gauche, C sur le plateau de droite.

Cette stratégie est illustrée par la matrice :

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

Le déterminant de la matrice $\mathbf{X}'\mathbf{X}$ est égal à :

$$|\mathbf{X}'\mathbf{X}| = 16$$

Cette stratégie est nettement meilleure que la méthode classique. Mais est-ce la meilleure ?

Pour les plans n’ayant que des niveaux -1 ou $+1$, la matrice $\mathbf{X}'\mathbf{X}$ de rang n ayant la plus haute D-optimalité possède un déterminant $|\mathbf{X}'\mathbf{X}| = n^n$.

Ainsi la seconde solution proposée n’est pas une stratégie optimale. Comme indiqué précédemment, les plans optimaux sont calculés par les ordinateurs. L’une des

méthodes la plus efficace pour calculer un plan optimal est l'algorithme d'échanges des coordonnées. Nous allons donner les grandes lignes de cette méthode et l'illustrer grâce à l'exemple de la pesée de quatre objets.

L'algorithme d'échanges des coordonnées suit la démarche suivante :

Préparation de l'algorithme :

- Création d'une matrice carrée X de rang n dont les éléments sont choisis au hasard et dont les valeurs sont comprises entre 0 et 1. C'est la matrice aléatoire de départ.
- Création d'un jeu de substitution dont les valeurs serviront à remplacer les éléments de départ. Ce jeu est constitué par les niveaux d'étude du plan d'expériences, ici -1 ou $+1$. Par exemple, pour un plan 2^2 , on prend -1 et $+1$ et, pour un plan à trois niveaux, on prend -1 , 0 et $+1$.

Ensuite commence le développement itératif de l'algorithme :

- Le chiffre à la croisée de la première ligne et de la première colonne est remplacé successivement par les valeurs du jeu de substitution. Le déterminant $|X'X|$ est calculé pour chacune des valeurs de substitution.
- Si l'une des valeurs de substitution augmente le déterminant, on la laisse en place en remplacement de la valeur d'origine. Si aucune des valeurs de substitution n'augmente le déterminant, on laisse la valeur aléatoire en place.
- Le premier élément étant traité, on attaque le second : il est à la croisée de la première ligne et de la deuxième colonne. On exécute les opérations 1 et 2 et on trouve une nouvelle valeur pour le déterminant $|X'X|$.
- On poursuit ainsi pour tous les éléments de la matrice X .
- Si l'on atteint la valeur maximale du déterminant, soit $|X'X| = n^n$, on s'arrête.
- Si l'on n'a pas atteint la valeur maximale, on répète les opérations de substitution sur tous les éléments de la matrice. On poursuit ainsi jusqu'à obtenir la valeur maximale du déterminant ou jusqu'à ce qu'on ne puisse plus améliorer sa valeur. Cette situation arrive lorsque l'algorithme ne procède plus à aucune substitution.

Reprenons l'exemple de la pesée de quatre objets pour illustrer ce processus et cherchons la meilleure matrice de pesées ou matrice d'expériences (Tableau 11.3).

Les éléments de la matrice de départ sont des nombres choisis au hasard :

$$X = \begin{bmatrix} 0,556 & 0,696 & 0,346 & 0,787 \\ 0,607 & 0,039 & 0,499 & 0,017 \\ 0,267 & 0,975 & 0,770 & 0,638 \\ 0,409 & 0,368 & 0,061 & 0,421 \end{bmatrix}$$

Dans ce cas :

$$|X'X| = 0,0016$$

Puisque les quatre facteurs possèdent deux niveaux, les valeurs candidates pour remplacer les éléments de cette matrice de départ sont -1 et $+1$.

Tableau 11.3 – Détail de l'algorithme d'échanges (pesées de 4 objets).

Itération	Matrice X	IX' Xi avec -1	IX' Xi avec +1	Valeur de substitution	Nouvelle matrice X
1	$X = \begin{bmatrix} 0,556 & 0,696 & 0,346 & 0,787 \\ 0,607 & 0,039 & 0,499 & 0,017 \\ 0,267 & 0,975 & 0,770 & 0,638 \\ 0,409 & 0,368 & 0,061 & 0,421 \end{bmatrix}$	0,027	2×10^{-5}	-1	$X = \begin{bmatrix} -1 & 0,696 & 0,346 & 0,787 \\ 0,607 & 0,039 & 0,499 & 0,017 \\ 0,267 & 0,975 & 0,770 & 0,638 \\ 0,409 & 0,368 & 0,061 & 0,421 \end{bmatrix}$
2	$X = \begin{bmatrix} -1 & 0,696 & 0,346 & 0,787 \\ 0,607 & 0,039 & 0,499 & 0,017 \\ 0,267 & 0,975 & 0,770 & 0,638 \\ 0,409 & 0,368 & 0,061 & 0,421 \end{bmatrix}$	0,332	0,008	-1	$X = \begin{bmatrix} -1 & -1 & 0,346 & 0,787 \\ 0,607 & 0,039 & 0,499 & 0,017 \\ 0,267 & 0,975 & 0,770 & 0,638 \\ 0,409 & 0,368 & 0,061 & 0,421 \end{bmatrix}$
3	$X = \begin{bmatrix} -1 & -1 & 0,346 & 0,787 \\ 0,607 & 0,039 & 0,499 & 0,017 \\ 0,267 & 0,975 & 0,770 & 0,638 \\ 0,409 & 0,368 & 0,061 & 0,421 \end{bmatrix}$	0,187	0,418	1	$X = \begin{bmatrix} -1 & -1 & 1 & 0,787 \\ 0,607 & 0,039 & 0,499 & 0,017 \\ 0,267 & 0,975 & 0,770 & 0,638 \\ 0,409 & 0,368 & 0,061 & 0,421 \end{bmatrix}$
4	$X = \begin{bmatrix} -1 & -1 & 1 & 0,787 \\ 0,607 & 0,039 & 0,499 & 0,017 \\ 0,267 & 0,975 & 0,770 & 0,638 \\ 0,409 & 0,368 & 0,061 & 0,421 \end{bmatrix}$	0,024	0,497	1	$X = \begin{bmatrix} -1 & -1 & 1 & 1 \\ 0,607 & 0,039 & 0,499 & 0,017 \\ 0,267 & 0,975 & 0,770 & 0,638 \\ 0,409 & 0,368 & 0,061 & 0,421 \end{bmatrix}$
...et ainsi de suite jusqu'aux deux dernières lignes :					
15	$X = \begin{bmatrix} -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 \\ 1 & 1 & 0,061 & 0,421 \end{bmatrix}$	32	187	1	$X = \begin{bmatrix} -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 0,421 \end{bmatrix}$
16	$X = \begin{bmatrix} -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 0,421 \end{bmatrix}$	64	256	1	$X = \begin{bmatrix} -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$

Le déterminant a atteint la valeur 256, soit 4^4 . Nous savons que c'est la valeur maximale. Par conséquent la matrice d'expériences optimale est la suivante :

$$X = \begin{bmatrix} -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

On constate que les quatre objets interviennent dans chaque pesée et c'est ce que nous attendions. La valeur de $|X'X| = 256$ indique que l'algorithme peut être arrêté. Si nous n'avions pas atteint la valeur 256 après l'échange des 16 éléments de la matrice, nous aurions parcouru à nouveau le processus de substitution avec, comme matrice de départ, la matrice d'arrivée du premier cycle. La puissance des ordinateurs permet d'affiner ce processus : on utilise plusieurs matrices de départ (une vingtaine), on parcourt plusieurs fois l'algorithme et on compare les matrices obtenues. On prend la meilleure.

Pour résumer, nous avons vu trois stratégies pour peser quatre objets. Les résultats sont les suivants :

- Stratégie classique $|X'X| = 1$
- Stratégie intermédiaire $|X'X| = 16$
- Stratégie optimale $|X'X| = 256$

11.4 Plans optimaux avec un modèle linéaire

On ne traite pas souvent des cas aussi simples que celui des pesées. En général, on a un problème qui met en jeu un modèle, une ou plusieurs réponses, plusieurs facteurs et leurs interactions. Comment les plans optimaux sont-ils calculés quand les modèles linéaires sont impliqués ?

Supposons que nous ayons le modèle linéaire suivant :

$$y = a_0 + a_1x_1 + a_2x_2 + a_{12}x_1x_2$$

On démarre avec deux facteurs et une interaction. Pour utiliser l'algorithme d'échanges précédent, on commence par construire la matrice du modèle. C'est cette matrice qui est désignée par X. Cette matrice se distingue de la matrice du plan d'expériences (repérée F dans ce chapitre) qui, elle, ne fait intervenir que les niveaux des facteurs. La matrice X fait intervenir les niveaux des facteurs, leurs interactions et une constante.

On commence par remplir la matrice d'expériences de nombres r_i choisis au hasard :

$$F = \begin{bmatrix} r_1 & r_2 \\ r_3 & r_4 \\ r_5 & r_6 \\ r_7 & r_8 \end{bmatrix}$$

On bâtit la matrice du modèle en ajoutant une colonne de 1 pour la constante et une quatrième colonne pour l'interaction :

$$\mathbf{X} = \begin{bmatrix} 1 & r_1 & r_2 & r_1 r_2 \\ 1 & r_3 & r_4 & r_3 r_4 \\ 1 & r_5 & r_6 & r_5 r_6 \\ 1 & r_7 & r_8 & r_7 r_8 \end{bmatrix}$$

L'algorithme d'échanges est appliqué à la matrice \mathbf{F} . Quand la matrice \mathbf{F} est modifiée, la matrice \mathbf{X} est modifiée de la même manière. Par exemple quand on remplace le r_1 de \mathbf{F} par $+1$ (ou -1), on modifie le r_1 de \mathbf{X} dans les deuxième et quatrième colonnes. Le déterminant correspondant est calculé, ce qui permet de savoir s'il faut conserver ou non le niveau testé.

Comme précédemment nous allons donner un exemple (Tableau 11.4). Soit \mathbf{F} la matrice d'expériences à optimiser :

$$\mathbf{F} = \begin{bmatrix} 0,736 & 0,414 \\ 0,239 & 0,124 \\ 0,328 & 0,770 \\ 0,318 & 0,723 \end{bmatrix}$$

D'où la matrice :

$$\mathbf{X} = \begin{bmatrix} 1 & 0,736 & 0,414 & 0,305 \\ 1 & 0,239 & 0,124 & 0,030 \\ 1 & 0,328 & 0,770 & 0,252 \\ 1 & 0,318 & 0,723 & 0,230 \end{bmatrix}$$

D'où la matrice finale :

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

Si l'on regarde attentivement cette matrice, on constate que les deux colonnes centrales forment un plan factoriel complet 2^2 , plan que nous avons rencontré au chapitre 2.

D'une manière générale, l'algorithme d'échanges permet de retrouver les plans classiques. C'est donc un moyen puissant de construire rapidement un plan qui s'adapte aux exigences de l'étude en cours même dans les cas simples.

Tableau 11.4 – Détail de l'algorithme d'échanges (plan de 2 facteurs).

Itération	F	IX' XI avec -1	IX' XI avec +1	Valeur de substitution	Nouvelle matrice X
1	$\mathbf{F} = \begin{bmatrix} 0,736 & 0,414 \\ 0,239 & 0,124 \\ 0,328 & 0,770 \\ 0,318 & 0,723 \end{bmatrix}$	$1,6 \times 10^{-5}$	$5,5 \times 10^{-5}$	-1	$\mathbf{X} = \begin{bmatrix} 1 & -1 & 0,414 & -0,414 \\ 1 & 0,239 & 0,124 & 0,030 \\ 1 & 0,328 & 0,770 & 0,252 \\ 1 & 0,318 & 0,723 & 0,230 \end{bmatrix}$
2	$\mathbf{F} = \begin{bmatrix} -1 & 0,414 \\ 0,239 & 0,124 \\ 0,328 & 0,770 \\ 0,318 & 0,723 \end{bmatrix}$	$5,6 \times 10^{-4}$	$6,6 \times 10^{-6}$	-1	$\mathbf{X} = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & 0,239 & 0,124 & 0,030 \\ 1 & 0,328 & 0,770 & 0,252 \\ 1 & 0,318 & 0,723 & 0,230 \end{bmatrix}$
...et ainsi de suite jusqu'aux deux dernières lignes :					
7	$\mathbf{F} = \begin{bmatrix} -1 & -1 \\ -1 & 1 \\ 1 & -1 \\ 0,318 & 0,723 \end{bmatrix}$	0	190	1	$\mathbf{X} = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 0,723 & 0,723 \end{bmatrix}$
8	$\mathbf{F} = \begin{bmatrix} -1 & -1 \\ -1 & 1 \\ 1 & -1 \\ 1 & 0,723 \end{bmatrix}$	0	256	1	$\mathbf{X} = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$

11.5 Quand utiliser des plans optimaux ?

Les propriétés des plans optimaux font qu'ils sont toujours un bon choix, qu'il s'agisse de criblage ou de modélisation. Ils permettent à l'expérimentateur d'avoir le plan sur mesure qui s'adapte parfaitement aux contraintes de l'étude. Les plans optimaux permettent également de traiter des situations difficiles et sont parfois la seule solution aux problèmes posés. On les emploie lorsqu'il faut :

- diminuer le nombre d'essais d'un plan classique pour s'adapter au budget et aux délais imposés ;
- adapter le plan aux objectifs de l'étude et aux moyens de l'expérimentateur ;
- traiter en même temps des facteurs de mélanges et des facteurs de procédé ;
- obtenir une surface de réponse en présence de facteurs discrets ;
- tenir compte des contraintes qui pèsent sur le domaine d'étude ; toutes les combinaisons de facteurs ne sont pas possibles et le domaine d'étude n'est plus un carré, un cube ou un hypercube ;
- assurer un blocking nécessaire à la qualité des résultats de l'étude et que le modèle ne s'y prête pas ;
- compléter par quelques essais supplémentaires une expérimentation mal conduite afin d'en tirer des conclusions valides (réparation d'expérimentation).

11.6 Adaptabilité des plans optimaux

Supposons que vous ayez une étude faisant intervenir une réponse et trois facteurs. Votre budget vous autorise 12 essais. Votre choix naturel se porte sur un plan 2^3 de 8 essais et il vous reste 4 essais que vous pouvez mettre au centre du domaine d'étude pour vérifier la validité du modèle PDAI (modèle du premier degré avec interactions) et pour avoir une idée de l'erreur expérimentale (Tableau 11.5).

Tableau 11.5 – Plan classique.

Facteur 1	Facteur 2	Facteur 3
-1	-1	-1
1	-1	-1
-1	1	-1
1	1	-1
-1	-1	1
1	-1	1
-1	1	1
1	1	1

Tableau 11.5 – Plan classique. (Suite)

Facteur 1	Facteur 2	Facteur 3
0	0	0
0	0	0
0	0	0
0	0	0

Si l'écart-type de la réponse est 1, l'écart-type de chaque coefficient est 0,125. On retrouve ce résultat avec la relation {5.1} du chapitre 5.

Si vous n'avez pas besoin de vérifier la validité du modèle PDAI et de posséder une estimation de l'erreur expérimentale, vous pouvez construire un plan optimal de 12 essais (Tableau 11.6).

Tableau 11.6 – Plan optimal pour 12 essais.

Facteur 1	Facteur 2	Facteur 3
-1	-1	-1
1	-1	-1
-1	1	-1
1	1	-1
-1	-1	1
-1	-1	1
1	-1	1
1	-1	1
-1	1	1
-1	1	1
1	1	1
1	1	1

Si l'écart-type de la réponse est 1, l'écart-type de chaque coefficient est 0,094. On retrouve aussi ce résultat avec la relation {5.1} du chapitre 5. La variance des coefficients chute de 0,125 à 0,094, soit un gain de près de 25 % sur les variances et de 14 % sur les écarts-types.

L'intérêt d'un tel gain réside dans le fait que les facteurs peu influents sont plus facilement discriminés de l'erreur expérimentale. Cet avantage peut être très important en fabrication où une légère amélioration du rendement de production peut parfois entraîner de grandes économies.

Si vous désirez établir, tout de suite, un modèle du second degré, il est possible en répartissant convenablement les 12 points d'expériences d'avoir une bonne précision sur les coefficients et d'estimer la courbure de la surface de réponse (Tableau 11.7). On choisira un modèle postulé du second degré :

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

Tableau 11.7 – Plan du second degré.

Facteur 1	Facteur 2	Facteur 3
1	-1	-1
-1	-1	1
-1	1	-1
0	1	0
1	1	1
-1	0	1
1	-1	1
-1	1	1
1	0	0
1	1	-1
-1	-1	-1
0	0	-1

Comme on utilise un modèle du second degré, la variance varie d'un coefficient à l'autre (Tableau 11.8).

Tableau 11.8 – Variance des coefficients du plan du second degré.

Coefficient	Variance	Coefficient	Variance
a_0	0,819	a_{13}	0,119
a_1	0,111	a_{23}	0,125
a_2	0,123	a_{11}	0,836
a_3	0,111	a_{22}	0,523
a_{12}	0,125	a_{33}	0,836

Le plan optimal a permis d'obtenir une variance faible et d'estimer les coefficients du second degré.

Avec le même nombre d'essais, il a été possible de traiter trois situations différentes : un plan classique avec vérification du modèle PDAI, un plan avec réduction maximale de la variance des coefficients et un plan pour surfaces de réponse.

Nous allons maintenant examiner trois applications différentes des plans optimaux en utilisant les trois exemples suivants :

- L'exemple 16 « Développement d'un détecteur de fissures » montre les propriétés d'adaptabilité des plans D-optimaux.
- L'exemple 17 « Les comprimés du pharmacien » souligne la réduction du nombre des essais dans le cas des études où se côtoient les facteurs de mélanges et les facteurs de procédé.
- L'exemple 18 « La crème à bronzer » témoigne de la puissance des plans D-optimaux pour réparer des expérimentations mal conçues.

11.7 Exemple 16 : développement d'un détecteur de fissures

11.7.1 Préparation du plan d'expériences

■ Présentation de l'étude

Un chercheur développe un appareil portable pour détecter les craquelures et les fissures dans les métaux. Cet appareil ne détruit pas l'objet examiné et se révèle très utile pour vérifier des produits comme les ailes d'avion, les roues de voiture, les matériaux de construction ou les canalisations. Le principe de ce détecteur est basé sur les modifications du champ magnétique d'un solénoïde dont on mesure les variations d'impédance. Au cours de l'étude ces variations seront appelées *sensibilité* de l'appareil, sensibilité qui sera la réponse mesurée par l'expérimentateur.

■ Facteurs et domaine d'étude

L'étude porte essentiellement sur l'enroulement des fils du solénoïde. Les trois facteurs étudiés sont (Tableau 11.9) :

- Facteur 1 : nombre d'enroulements des fils.
- Facteur 2 : distance des fils entre deux enroulements.
- Facteur 3 : calibre des fils.

Tableau 11.9 – Facteurs et domaine d'étude.

Facteur	Niveau -1	Niveau 0	Niveau +1
Nombre d'enroulements (1)	90	135	180
Distance des enroulements (2)	0,38	0,76	1,14
Calibre des fils (3)	40	44	48

■ Objectifs de l'étude

Pour cette étude, on peut réaliser 12 essais. Nous allons considérer deux expérimentateurs ayant des objectifs différents. Selon l'objectif de l'étude, on ne réalisera pas le même plan d'expériences. Deux plans vont être réalisés ayant chacun un objectif bien précis mais différent dans les deux cas. Cet exemple va nous permettre de comprendre l'intérêt de bien préciser l'objectif de l'étude avant de faire les essais. Il nous montrera aussi l'intérêt des plans classiques et des plans optimaux pour adapter le plan d'expériences aux impératifs d'une étude.

Le premier expérimentateur a pour objectif de modéliser le phénomène qu'il étudie. Il choisit donc un plan factoriel complet et dispose les quatre essais supplémentaires au centre du domaine d'étude pour apprécier la courbure de la surface de réponse et l'erreur expérimentale. Si besoin est, il complétera ce plan par des points en étoile pour avoir un plan composite.

Le second expérimentateur a pour objectif de détecter les facteurs influents et même ceux qui sont peu influents. Dans ce cas, il prévoit de faire un plan de criblage en utilisant les quatre essais supplémentaires pour diminuer la variance des coefficients. Il pourra ainsi mieux apprécier si un facteur est négligeable ou non.

11.7.2 Plan factoriel 2^3 avec points au centre

■ Expérimentation

Dans ce premier cas, l'objectif de l'expérimentateur est de modéliser le phénomène qu'il étudie. Il commence donc par un plan factoriel complet avec 4 points de contrôle au centre (Tableau 11.10).

Tableau 11.10 – Matrice d'expériences.

Essai n°	Nombre d'enroulements (1)	Distance des enroulements (2)	Calibre des fils (3)	Sensibilité
1	-1	-1	-1	1,7
2	+1	-1	-1	4,57
3	-1	+1	-1	0,55
4	+1	+1	-1	3,39
5	-1	-1	+1	1,51
6	+1	-1	+1	4,59
7	-1	+1	+1	0,67
8	+1	+1	+1	4,29
9	0	0	0	2,7
10	0	0	0	2,76

Tableau 11.10 – Matrice d'expériences. (Suite)

Essai n°	Nombre d'enroulements (1)	Distance des enroulements (2)	Calibre des fils (3)	Sensibilité
11	0	0	0	3,21
12	0	0	0	1,97
Niveau -1	90	0,38	40	
Niveau +1	180	1,14	48	

■ Interprétation

Les 8 coefficients sont calculés avec le modèle A en considérant les 8 points du plan factoriel et sans prendre en compte les points de contrôle :

Modèle A :

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

La constante du modèle est égale à 2,659 et il n'y a aucune indication statistique puisque le plan est saturé. Les points au centre ont pour moyenne 2,66 avec un écart-type de 0,51. La constante (réponse calculée au centre avec le modèle) et les réponses mesurées au centre du domaine d'étude sont statistiquement égales. Le modèle PDAI est donc valide, ce n'est pas la peine d'envisager un plan composite. Comme on sait maintenant que les points au centre font partie du modèle PDAI, on peut les incorporer aux calculs des coefficients.

On reprend donc les calculs en conservant le modèle PDAI mais en prenant en compte les 12 résultats expérimentaux. Le tableau 11.11 indique les résultats de ce calcul. La constante est égale à 2,659 et il y a, cette fois, des indications statistiques. Le facteur 1 (nombre d'enroulements) est influent et le facteur 2 est peut-être influent mais la conclusion n'est pas nette car la *p*-value est de 0,05.

Tableau 11.11 – Importance des coefficients du modèle A calculés avec les points au centre (unités codées).

Coefficient	Valeur	Écart-type	t de Student	p-value
Constante	2,659	0,128	20,73	< 0,0001
Nombre d'enroulements (1)	1,551	0,157	9,87	0,0006
Distance des enroulements (2)	-0,434	0,157	-2,76	0,0508
Calibre fils (3)	0,106	0,157	0,68	0,5360
a_{12}	0,064	0,157	0,41	0,7057
a_{13}	0,124	0,157	0,79	0,4750
a_{23}	0,149	0,157	0,95	0,3974
a_{123}	0,071	0,157	0,45	0,6738

En examinant les résultats, on s'aperçoit que les interactions sont négligeables. On refait les calculs avec un modèle simplifié, le modèle B, ne faisant intervenir que les facteurs principaux au premier degré :

$$\text{Modèle B : } y = a_0 + a_1x_1 + a_2x_2 + a_3x_3$$

On gagne ainsi des degrés de liberté qui permettent de réduire l'erreur sur les coefficients. On constate que le coefficient du facteur 2 (distance des enroulements) a une p -value de 0,0123 (Tableau 11.12). Ce facteur est donc influent.

Tableau 11.12 – Importance des coefficients du modèle B (unités codées).

Coefficient	Valeur	Écart-type	t de Student	p-value
Constante	2,659	0,110	24,16	< 0,0001
Nombre d'enroulements (1)	1,551	0,135	11,51	< 0,0001
Distance des enroulements (2)	-0,434	0,135	-3,22	0,0123
Calibre fils (3)	0,106	0,135	0,79	0,4533

On peut conclure que le modèle est du premier degré et qu'il y a deux facteurs influents : le facteur 1, le nombre d'enroulements, et le facteur 2, la distance des enroulements.

L'objectif de modélisation du phénomène est atteint.

11.7.3 Plan D-optimal avec répétitions

Dans ce second cas, l'objectif de l'expérimentateur est de détecter les facteurs peu influents. Il va donc faire un plan de criblage en utilisant un modèle PDAI et en essayant de diminuer le plus possible la variance des coefficients pour bien apprécier leur signification. Pour cela il met tous les points sur les sommets du domaine d'étude. Il y aura donc des répétitions sur certains sommets.

■ Expérimentation

Le plan d'expériences est calculé par le logiciel en demandant un plan D-optimal. Les résultats des essais figurent dans le tableau 11.13.

Tableau 11.13 – Matrice d'expériences.

Essai n°	Nombre d'enroulements (1)	Distance des enroulements (2)	Calibre des fils (3)	Sensibilité
1	-1	-1	-1	1,190
2	-1	-1	-1	2,210
3	-1	-1	1	1,510
4	-1	1	-1	0,823

Tableau 11.13 – Matrice d'expériences. (Suite)

Essai n°	Nombre d'enroulements (1)	Distance des enroulements (2)	Calibre des fils (3)	Sensibilité
5	-1	1	-1	0,277
6	-1	1	1	0,670
7	1	-1	-1	4,822
8	1	-1	-1	4,318
9	1	-1	1	4,590
10	1	1	-1	3,402
11	1	1	-1	3,378
12	1	1	1	4,290
Niveau -1	90	0,38	40	
Niveau +1	180	1,14	48	

■ Interprétation

Les coefficients sont calculés avec le modèle A et leurs propriétés statistiques sont indiquées dans le tableau 11.14.

Tableau 11.14 – Importance des coefficients du modèle A (unités codées).

Coefficient	Valeur	Écart-type	t de Student	p-value
Constante	2,659	0,137	19,46	< 0,0001
Nombre d'enroulements (1)	1,551	0,137	11,35	0,0003
Distance des enroulements (2)	-0,434	0,137	-3,17	0,0337
Calibre des fils (3)	0,106	0,137	0,78	0,4802
a_{12}	0,064	0,137	0,47	0,6651
a_{13}	0,124	0,137	0,91	0,4163
a_{23}	0,149	0,137	1,09	0,3375
a_{123}	0,071	0,137	0,52	0,6296

Les résultats montrent qu'il y a deux facteurs influents, le facteur 1 (nombre d'enroulements) et le facteur 2 (distance des enroulements). Comme précédemment, on constate que les interactions sont négligeables et que l'on peut reprendre les calculs en utilisant le modèle B. Les résultats de ce dernier calcul figurent dans le tableau 11.15 et confirment l'influence du facteur 2 qui a maintenant une p -value de 0,003.

Tableau 11.15 – Importance des coefficients du modèle B (unités codées).

Coefficient	Valeur	Écart-type	t de Student	p -value
Constante	2,659	0,122	21,80	< 0,0001
Nombre d'enroulements (1)	1,51	0,115	13,13	< 0,0001
Distance des enroulements (2)	-0,483	0,115	-4,20	0,0030
Calibre des fils (3)	0,106	0,122	0,87	0,4090

Tous les essais du plan ont été utilisés pour réduire l'écart-type des coefficients permettant ainsi de découvrir même les facteurs peu influents.

11.8 Exemple 17 : les comprimés du pharmacien

Ce plan d'expériences a été décrit dans le livre *Plans d'expériences : les mélanges*. En ce qui concerne les facteurs de mélanges, les expérimentateurs avaient décidé de réaliser 11 essais localisés comme l'indique la figure 11.4. En ce qui concerne les facteurs de procédé, les expérimentateurs avaient décidé de réaliser un plan factoriel complet avec point central. L'ensemble de l'expérimentation avait conduit à 11 fois 5 soit 55 essais. Grâce aux plans D-optimaux, il est possible de réduire considérablement le nombre de ces essais tout en assurant une modélisation d'aussi bonne qualité. Nous allons rappeler les données de l'étude et nous comparerons les résultats obtenus avec 55 essais à ceux obtenus avec le plan D-optimal.

11.8.1 Préparation du plan d'expériences

■ Présentation de l'étude

Cette étude concerne la mise au point de cachets pharmaceutiques. Les expérimentateurs ont cherché à atteindre plusieurs spécifications en étudiant la composition du cachet et la manière de l'obtenir. Le cachet est préparé à partir d'une poudre contenant de l'amidon, du lactose et de la cellulose. Cette poudre est soumise à une pression importante pendant un laps de temps déterminé. La poudre ainsi traitée devient sous l'effet de la pression, un comprimé. Le mot comprimé sera, dans la suite de cet exemple, synonyme de cachet. Le cachet ou le comprimé ainsi obtenu doit répondre à plusieurs impératifs pour pouvoir être utilisé par le patient.

Il doit pouvoir être tenu avec les doigts sans se désagréger, ne pas s'effriter dans la boîte lors des manipulations, avoir une porosité donnée pour se dissoudre facilement et rapidement, etc. L'étude de plusieurs mélanges amidon-lactose-cellulose en fonction de la pression et de la durée de compression a pour objectif de déterminer les conditions d'obtention de comprimés conformes aux spécifications.

■ Données du problème

□ Réponses

Plusieurs réponses définissent les qualités d'un cachet. Nous n'en utiliserons que deux pour traiter cet exemple : la cohésion et la porosité.

• Cohésion

Cette grandeur correspond à la résistance à la rupture. Elle est mesurée en exerçant une traction sur le cachet et en mesurant la force nécessaire pour le rompre. Cette force ne doit pas être trop faible, car sinon les cachets se casseraient sous la moindre traction. Elle ne doit pas, non plus, être trop forte, car il faut que le patient puisse facilement partager le comprimé en deux. On recherche une cohésion comprise entre deux limites : 3×10^6 et 5×10^6 Pa.

• Porosité

Cette réponse mesure le volume des vides du cachet. Elle s'obtient en comparant le volume de la poudre initiale à celui du cachet compressé. La porosité est donnée par la diminution relative de volume lorsque la poudre devient un comprimé. L'eau doit pouvoir pénétrer facilement à l'intérieur du comprimé pour le dissoudre rapidement, par conséquent la porosité ne doit pas être trop faible. Pour que le cachet ne s'effrite pas, la porosité ne doit pas être trop grande. Un bon compromis est obtenu lorsque la porosité est comprise entre 17 et 22 %.

□ Facteurs de mélanges

Le mélange à étudier comprend trois constituants :

– l'amidon dont la teneur est x_1 ; la teneur de ce produit est soumise à une contrainte : pas plus de 50 % d'amidon dans la poudre à cachets :

$$x_1 \leq 0,50$$

– le lactose dont la teneur x_2 n'est soumise à aucune contrainte :

$$0 \leq x_2 \leq 1$$

– la cellulose minérale dont la teneur x_3 n'est soumise à aucune contrainte :

$$0 \leq x_3 \leq 1$$

□ Facteurs de procédé

• Cadence de fabrication

C'est le temps qui s'écoule entre la compression d'un cachet et celle du suivant. Ce temps est en relation directe avec la durée de la pression exercée sur le cachet. On

remarquera également que ce facteur a une forte répercussion sur la production. En effet, plus la durée de compression est courte, plus la production est élevée. Nous désignerons ce facteur par z_1 .

- Pression

C'est la pression appliquée sur la poudre au moment de la fabrication du comprimé. Nous désignerons ce facteur par z_2 .

□ **Domaine d'étude**

Il faut tenir compte de la contrainte pour définir le domaine d'étude. Seule la partie inférieure du triangle équilatéral fait partie du domaine d'étude (Figure 11.4).

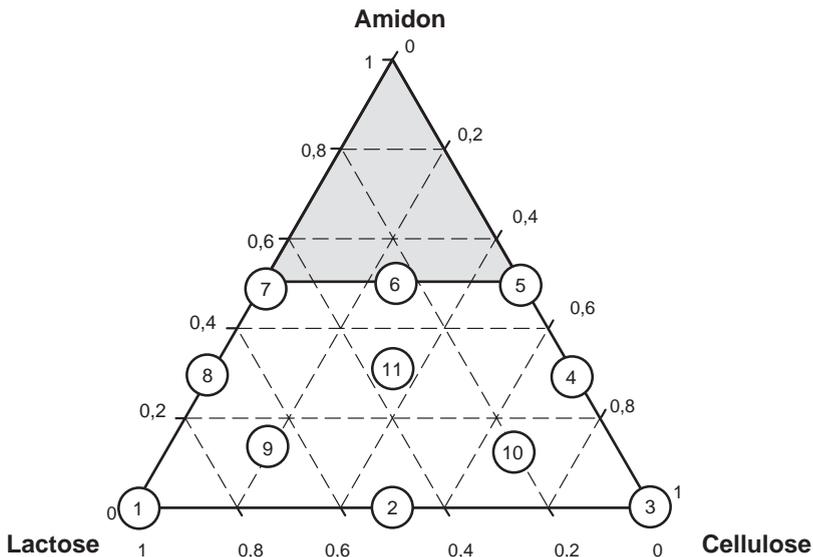


Figure 11.4 – Domaine d'étude des mélanges de poudres pour fabriquer des comprimés. Position des mélanges réalisés avec le plan de 55 essais.

11.8.2 Plan d'expériences

Pour construire le plan, il faut d'abord définir le modèle postulé. On peut, par exemple, prendre le modèle du second degré des plans de mélange :

$$y = b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3$$

Les coefficients de ce modèle dépendent des niveaux des facteurs de procédé z_1 et z_2 . On suppose que les coefficients sont des fonctions du premier degré des facteurs z_1 et z_2 :

$$b_i = a_{0,i} + a_{1,i}z_1 + a_{2,i}z_2$$

Avec ces choix, le modèle postulé s'écrit :

$$y = (a_{0,1} + a_{1,1}z_1 + a_{2,1}z_2)x_1 + (a_{0,2} + a_{1,2}z_1 + a_{2,2}z_2)x_2 + (a_{0,3} + a_{1,3}z_1 + a_{2,3}z_2)x_3 \\ + (a_{0,12} + a_{1,12}z_1 + a_{2,12}z_2)x_1x_2 + (a_{0,13} + a_{1,13}z_1 + a_{2,13}z_2)x_1x_3 \\ + (a_{0,23} + a_{1,23}z_1 + a_{2,23}z_2)x_2x_3$$

Ce modèle comporte 18 coefficients. Il faut donc au moins 18 essais différents pour l'établir. Mais on désire pouvoir s'assurer de la qualité du modèle et on introduit 10 essais supplémentaires. On va réaliser 28 essais, soit presque la moitié des essais du plan initial de 55 essais. Ce plan de 28 essais ne peut pas être construit à la main, il faut utiliser un logiciel possédant un module de calcul spécifique. L'ordinateur exécute les calculs et fournit l'un des plans optimaux possibles.

11.8.3 Expérimentation

Le tableau 11.6 indique le plan d'expériences qui a été retenu par les expérimentateurs.

Tableau 11.16 – Plan d'expériences.

Essai n°	Amidon (x_1)	Lactose (x_2)	Cellulose (x_3)	Cadence (z_1)	Pression (z_2)	Cohésion	Porosité
1	0,5	0	0,5	1	1	1,12	19,43
2	0,5	0,25	0,25	0	0	0,47	21,60
3	0	0	1	-1	-1	2,52	25,45
4	0	1	0	1	1	3,46	10,60
5	0,5	0,5	0	-1	1	0,49	15,88
6	0,5	0,5	0	1	-1	0,42	26,28
7	0	0	1	1	-1	2,27	27,96
8	0,25	0,35	0,4	-1	-1	0,51	26,19
9	0,25	0,375	0,375	1	1	1,80	14,56
10	0	1	0	1	-1	0,42	26,05
11	0,25	0	0,75	1	-1	0,31	28,10
12	0,25	0	0,75	0	1	2,87	14,94
13	0	0,5	0,5	-1	1	5,25	12,66
14	0,25	0,75	0	0	-1	0,19	24,83
15	0,5	0,5	0	1	1	0,38	18,20
16	0	1	0	-1	-1	0,91	21,62
17	0,5	0,5	0	-1	-1	0,53	21,82
18	0	1	0	-1	1	3,12	9,01
19	0,5	0	0,5	1	-1	0,41	27,81
20	0	0,5	0,5	-1	-1	1,49	24,37

Tableau 11.16 – Plan d'expériences. (Suite)

Essai n°	Amidon (x_1)	Lactose (x_2)	Cellulose (x_3)	Cadence (z_1)	Pression (z_2)	Cohésion	Porosité
21	0,5	0	0,5	-1	-1	0,07	25,83
22	0,5	0	0,5	-1	1	1,16	16,93
23	0	0,5	0,5	1	1	5,35	9,59
24	0	0	1	-1	1	7,09	8,32
25	0	0	1	1	1	7,39	10,65
26	0,25	0,75	0	-1	0	0,34	18,49
27	0	0,5	0,5	1	-1	1,49	23,29
28	0,25	0,75	0	1	0	0,72	20,55

11.8.4 Interprétation

Les calculs sont effectués comme à l'accoutumée et l'on obtient les coefficients des modèles Cohésion et Porosité. Ces coefficients peuvent être comparés à ceux qui ont été obtenus avec le plan de 55 essais. On constate que, pour les deux réponses, les coefficients sont statistiquement égaux (Tableau 11.17). On peut aussi comparer les écarts-types (RMSE) sur les réponses individuelles : 0,223 pour la cohésion et 0,654 pour la porosité. Ils étaient de 0,256 pour la cohésion et de 0,448 pour la porosité avec les modèles établis avec 55 essais. On obtient donc, là aussi, des résultats tout à fait semblables d'un point de vue statistique.

Avec les plans optimaux il est donc possible de réduire de manière très significative le nombre d'essais sans perdre les qualités statistiques des modèles établis.

Tableau 11.17 – Comparaison des coefficients des modèles Cohésion et Porosité établis avec 55 ou 28 essais (unités codées).

Coefficient	Cohésion 55	Cohésion 28	Porosité 55	Porosité 28
a_{01}	4,84	4,81	18,04	18,35
a_{02}	1,99	1,99	16,84	16,84
a_{03}	4,80	4,80	18,10	18,09
$a_{0,12}$	-11,80	-11,71	12,57	11,93
$a_{0,13}$	-16,69	-16,61	17,66	17,05
$a_{0,23}$	-0,01	0,03	0,11	-0,03
a_{11}	0,05	0,12	7,34	7,15

Tableau 11.17 – Comparaison des coefficients des modèles Cohésion et Porosité établis avec 55 ou 28 essais (unités codées). (Suite)

Coefficient	Cohésion 55	Cohésion 28	Porosité 55	Porosité 28
a_{12}	0,04	0,00	1,54	1,52
a_{13}	0,00	0,00	1,20	1,20
$a_{1,12}$	-0,10	-0,24	-10,90	-10,47
$a_{1,13}$	-0,03	-0,18	-12,61	-12,30
$a_{1,23}$	-0,01	-0,01	-9,66	-9,61
a_{21}	0,94	0,89	-0,16	0,07
a_{22}	1,31	1,31	-7,04	-7,02
a_{23}	2,45	2,45	-8,60	-8,60
$a_{2,12}$	-4,67	-4,58	0,34	-0,20
$a_{2,13}$	-4,86	-4,78	0,217	-0,14
$a_{2,23}$	-0,01	-0,02	5,87	5,80

On peut illustrer le tableau 11.17 par des diagrammes à barres qui révèlent immédiatement la concordance des résultats (Figures 11.5 et 11.6). Les conclusions de l'étude avec 28 essais seront donc les mêmes que celles obtenues avec 55 essais.

Cohésion : comparaison des coefficients des modèles établis avec 55 ou 28 essais

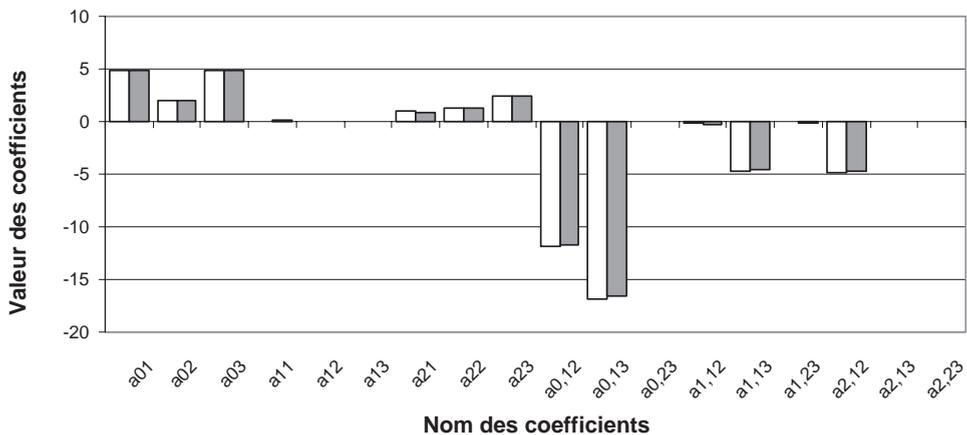


Figure 11.5 – Diagramme à barres comparant les coefficients des modèles Cohésion obtenus avec 55 essais (en blanc) et avec 28 essais (en gris).

Porosité : comparaison des coefficients des modèles établis avec 55 ou 28 essais

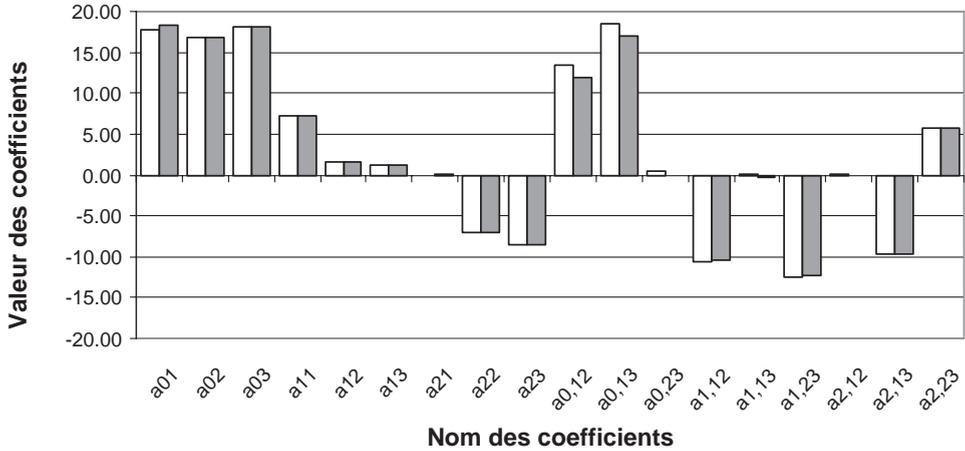


Figure 11.6 – Diagramme à barres comparant les coefficients des modèles Porosité obtenus avec 55 essais (en blanc) et avec 28 essais (en gris).

11.8.5 Conclusion de l'étude

Grâce aux modèles, on peut tracer les diagrammes (cadence × pression) pour différentes compositions du mélange amidon-lactose-cellulose. On s'aperçoit que la porosité peut être comprise entre 17 et 22 pour toutes les compositions du mélange à condition de bien choisir les niveaux de la cadence et de la pression (Figure 11.7).

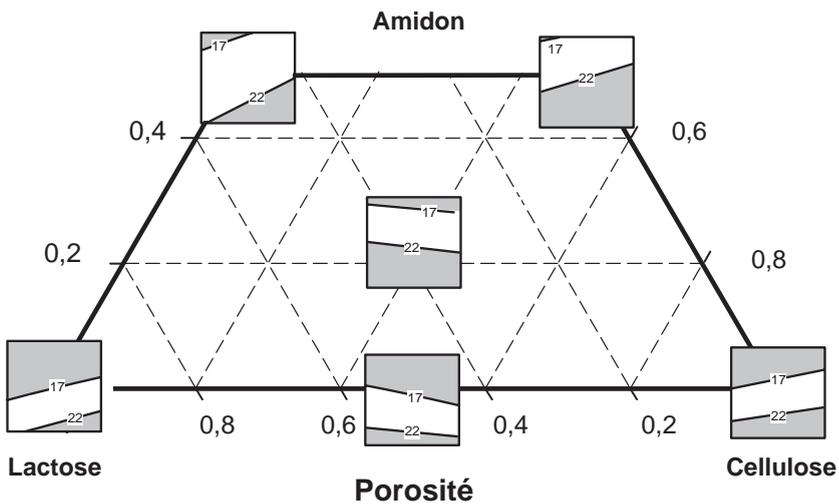


Figure 11.7 – L'objectif de porosité (partie blanche) des diagrammes (cadence × pression) peut être atteint pour toutes les compositions du mélange.

On constate également qu'une cohésion comprise entre 3 et 5 ne peut être respectée que pour les faibles teneurs en amidon (Figure 11.8).

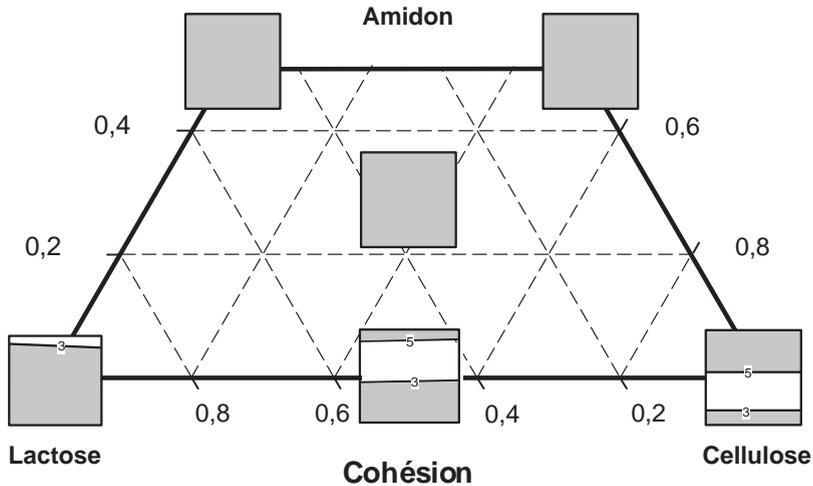


Figure 11.8 – L'objectif de cohésion (partie blanche) des diagrammes (cadence \times pression) ne peut être atteint que pour les basses teneurs en amidon.

Plusieurs solutions sont possibles et l'expérimentateur choisira celles qui, par exemple, diminuent les coûts de fabrication en travaillant à cadence élevée. Plusieurs mélanges caractérisés par une forte teneur en cellulose sont possibles pour maximiser la production. Par exemple, un mélange contenant 94 % de cellulose, 4 % de lactose et 2 % d'amidon pourrait convenir. Un mélange contenant 80 % de cellulose, 12 % de lactose et 8 % d'amidon conviendrait également. L'expérimentateur choisit de vérifier les conclusions de l'interprétation du plan d'expériences avec le mélange : 85 % de cellulose, 10 % de lactose et 5 % d'amidon.

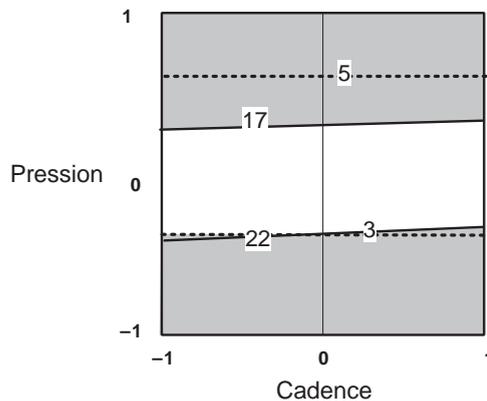


Figure 11.9 – Avec le mélange (85 % de cellulose + 10 % de lactose + 5 % d'amidon), les objectifs sont atteints si la pression reste entre les niveaux +0,15 et -0,40.

Les expériences de vérification confirment les prévisions et l'expérimentateur peut faire ses recommandations en donnant la composition du mélange et les réglages des facteurs de procédé. La cadence peut prendre tous les niveaux entre -1 et $+1$ mais la pression doit être contrôlée et rester entre les niveaux $+0,15$ et $-0,40$.

11.9 Exemple 18 : la crème à bronzer

Les plans optimaux sont également très utiles pour sauver un plan mal conçu. Cette situation apparaît quand l'expérimentateur n'organise pas ses essais convenablement. Les points d'expériences peuvent être dispersés de manière aléatoire ou suivre une logique incompatible avec l'obtention d'un modèle permettant des prévisions fiables. Citons un cas extrême pour bien faire ressortir les conséquences d'une mauvaise organisation des essais : on suppose que les points sont disposés comme l'indique la figure 11.10.

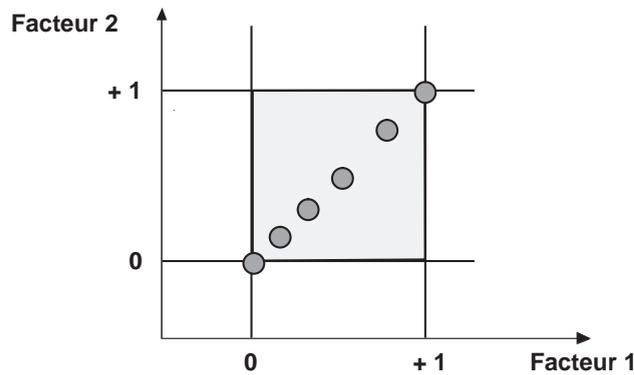


Figure 11.10 – Disposition des points d'expériences.

L'expérimentateur est intéressé par un modèle du premier degré :

$$y = a_0 + a_1x_1 + a_2x_2$$

Mais cette disposition des points d'expériences pose un gros problème. En effet, les niveaux des deux facteurs sont liés par la relation :

$$\frac{x_1}{x_2} = 1$$

Le modèle devient :

$$y = a_0 + (a_1 + a_2)x_1$$

On ne peut estimer que la somme des coefficients.

Cet exemple caricatural montre les conséquences d'une corrélation, c'est-à-dire d'une relation linéaire, entre les niveaux des facteurs. Si les points d'expériences sont alignés ou presque alignés, il y a une forte corrélation entre les niveaux des facteurs et les coefficients sont très fortement dépendants les uns des autres et possèdent des écarts-types très élevés. Les modèles sont alors inutilisables pour faire des prévisions précises puisque les incertitudes sur les coefficients se transmettent aux réponses calculées. Faut-il pour autant ne pas utiliser les résultats déjà obtenus ?

Cette introduction nous permettra de mieux comprendre l'exemple suivant sur la mise au point d'une crème de bronzage.

11.9.1 Préparation du plan d'expériences

■ Présentation de l'étude

Les crèmes à bronzer sont préparées à partir d'un sucre naturel dont on extrait un principe actif permettant de pigmenter la peau. Le principe actif est la dihydroacéto-
tone (DHA) qui est préparée industriellement par un procédé de cristallisation. La DHA brute est dissoute dans un mélange eau-solvant organique. Ce mélange est refroidi lentement de 40 °C à -10 °C sous agitation. Pendant le refroidissement, les cristaux se forment et sont récupérés par filtration. Le responsable de la fabrication avait constaté que le rendement de fabrication était compris entre 40 et 48 %. Il aurait bien aimé augmenter ce rendement. Il pouvait modifier le rapport eau/solvant et la vitesse d'agitation qui étaient respectivement égaux, dans les conditions normales de fabrication, à 70 % et à 200 tr/min.

Pour cela il décida de faire quelques essais en modifiant le rapport eau/solvant organique entre 60 et 80 % et la vitesse d'agitation entre 160 et 240 tr/min. Il effectua quatre essais différents et prit la précaution de faire deux expériences pour chaque essai.

■ Facteurs et domaine

L'expérimentateur retient donc les facteurs et les domaines suivants :

- Facteur 1 : rapport eau/solvant organique (entre 60 et 80 %).
- Facteur 2 : agitation (entre 160 et 240 tr/min).

Il réalise les essais et mesure avec soin les rendements obtenus. Pour être sûr d'avoir une expérimentation complète, il ajoute deux expériences réalisées dans les conditions normales de fabrication.

11.9.2 Expérimentation

Les résultats expérimentaux sont consignés dans le tableau 11.18.

Tableau 11.18 – Matrice d'expériences (unités non codées) et résultats expérimentaux.

Essai n°	Rapport eau/solvant organique (% d'eau)	Agitation (tr/min)	Rendement (%)	
0	70	200	45,5	43
1	80	240	44	39,5
2	75	224	41	40
3	61	160	46	49
4	60	160	45	49,5

11.9.3 Interprétation

■ Première interprétation

L'expérimentateur commence par calculer les coefficients du modèle polynomial et leur écart-type en utilisant les unités codées. À sa grande déception aucun des coefficients n'est significatif. Que se passe-t-il donc et pourquoi ces expériences n'apportent-elles pas une réponse à la question posée ? Le rendement serait-il toujours à peu près le même malgré les changements de conditions opératoires ?

Tableau 11.19 – Importance des coefficients du modèle de la réponse rendement calculée avec le modèle 1 (unités codées).

Coefficient	Valeur	Écart-type	t de Student	p-value
Constante	41,875	3,716	11,27	0,0004
Rapport eau/solvant organique (1)	5,125	21,86	0,23	0,8261
Agitation (2)	-7,875	21,99	-0,37	0,7291
a_{12}	2,625	4,48	0,59	0,5898

Pour essayer de comprendre, l'expérimentateur écrit le modèle :

$$y = 47,9 + 5,1 x_1 - 7,9 x_2 + 2,6 x_1 x_2 \quad \{\text{modèle 1}\}$$

puis il calcule les réponses prévues et regarde les intervalles de confiance à 95 %. Ces intervalles de confiance sont grands et il est statistiquement impossible de distinguer les réponses les unes des autres.

Tableau 11.20 – Intervalles de confiance des réponses prévues (modèle 1).

Essai n°	Rapport eau/solvant organique (% d'eau)	Agitation (tr/min)	Rendement prévu	Intervalle de confiance à 95 %
1	80	240	41,7	33,2 à 50,3
2	75	224	40,5	31,9 à 49
3	61	160	47,5	38,9 à 56,0
4	60	160	47,25	38,7 à 55,8

Troublé, l'expérimentateur décide d'aller voir l'expert en plans d'expériences de son entreprise.

■ Consultation de l'expert en plans d'expériences

L'expert commence par regarder où se localisent les expériences dans le domaine expérimental. Il trace un diagramme analogue à la figure 11.11. Par un hasard tout

à fait malheureux, les points d'expériences sont presque alignés. La surface délimitée par les points d'expériences est toute petite. C'est là que se trouve l'origine de tous les ennuis de l'expérimentateur.

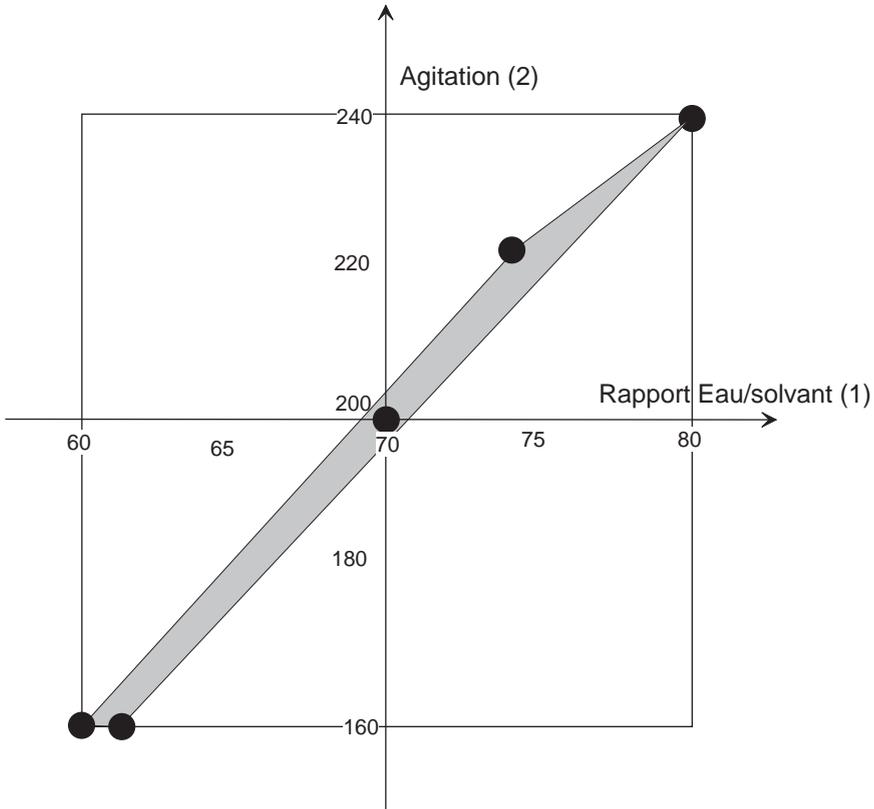


Figure 11.11 – Emplacement des points d'expériences et domaine expérimental.

L'expert explique alors à l'expérimentateur que le problème est facile à corriger. Il suffit d'agrandir le domaine d'étude en incorporant des points d'expériences supplémentaires bien placés. Dans le cas de deux facteurs, on peut réaliser cette opération facilement. Lorsqu'il y a plus de deux facteurs, l'opération ne peut se faire qu'avec un logiciel qui sait comment agrandir un volume multidimensionnel avec le minimum de points d'expériences : le logiciel maximise le déterminant de la matrice $X'X$. La réparation du plan d'expériences consiste donc à construire un plan D-optimal en ajoutant des points d'expériences aux essais déjà réalisés. L'expert continue d'expliquer qu'avec deux facteurs il est facile de voir si les points d'expériences sont alignés ou presque alignés mais que pour les espaces multidimensionnels, il faut utiliser les corrélations entre les niveaux des facteurs. Si la corrélation entre les niveaux de deux facteurs est proche de 1, les points représentatifs sont alignés. Au fur et à mesure que la corrélation diminue, les points sont de moins en moins alignés et, lorsque la corrélation est nulle, les points ne sont plus

du tout alignés. Les plans factoriels, les plans pour surfaces de réponse classiques et les plans D-optimaux répondent à ces impératifs.

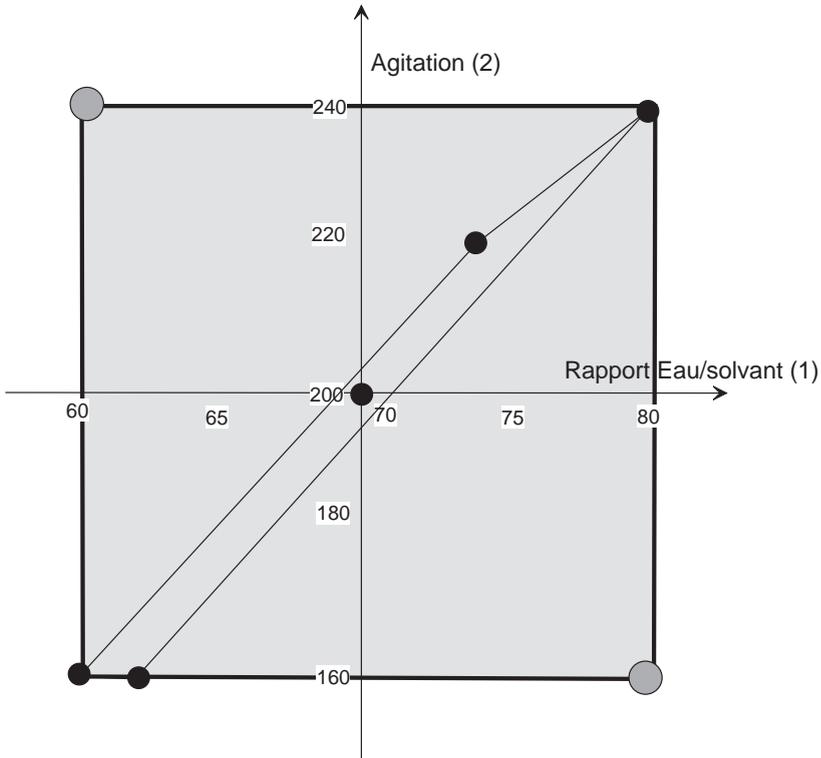


Figure 11.12 – Emplacement des points d'expériences supplémentaires et nouveau domaine expérimental.

■ Essais réparateurs et nouvelle interprétation

Il faut, expliqua l'expert, utiliser tout le domaine expérimental. Puisque vous aviez décidé de faire varier le rapport eau/solvant organique entre 60 et 80 %, il faut réaliser les expériences à ces valeurs. Il en est de même pour la vitesse d'agitation qui doit être testée à 160 et à 240 tr/min. Il suffit donc d'ajouter deux essais à ceux que vous avez déjà faits et vous devriez, cette fois, obtenir de bons résultats. Ces essais sont indiqués dans le tableau 11.21 et illustrés par la figure 11.12.

Tableau 11.21 – Les essais réparateurs et leurs résultats expérimentaux (unités non codées).

Essai n°	Rapport eau/solvant organique (% d'eau)	Agitation (tr/min)	Rendement (%)	
5	60	240	59,5	60
6	80	160	24	21

L'expérimentateur réalise alors les deux essais supplémentaires et les répète pour être homogène avec la première série d'essais. Il reprend les calculs en incorporant tous les essais réalisés sauf les points de contrôle au centre. Puis, il consigne ses résultats dans un tableau (Tableau 11.22) et écrit le modèle correspondant :

$$y = 42,6 - 11,1 x_1 + 7,5 x_2 + 1,7 x_1 x_2 \quad \{\text{modèle 2}\}$$

Tableau 11.22 – Importance des coefficients du modèle 2 sur la réponse « Rendement » (unités codées).

Coefficient	Valeur	Écart-type	t de Student	p-value
Constante	42,62	0,659	64,69	< 0,0001
Rapport eau/solvant organique (1)	-11,07	0,722	-15,33	< 0,0001
Agitation (2)	7,49	0,702	10,66	< 0,0001
a_{12}	1,75	0,737	2,38	0,0447

L'expérimentateur utilise le modèle 2 pour calculer les réponses prévues et regarde les intervalles de confiance. Ces intervalles sont plus petits que précédemment et la différence entre la plus faible et la plus forte valeur est significative. Il sait maintenant qu'il peut améliorer le rendement des cristaux de DHA en diminuant le rapport eau/solvant organique et en augmentant la vitesse d'agitation. Il peut espérer un rendement de 60 % avec un intervalle de confiance à 95 % de plus ou moins 7 %.

Tableau 11.23 – Intervalles de confiance des réponses prévues avec le modèle 2.

Essai n°	Rapport eau/solvant organique (% d'eau)	Agitation (tr/min)	Rendement prévu	Intervalle de confiance à 95 %
1	80	240	40,78	34,8 à 46,8
2	75	224	42,10	36,6 à 47,6
3	61	160	46,7	41 à 52,4
4	60	160	47,9	42,2 à 53,7
5	60	240	59,4	53,2 à 65,7
6	60	160	22,3	16 à 28,5

L'expérimentateur procède alors aux essais de vérification et constate que les prévisions sont bien en accord avec les résultats expérimentaux.

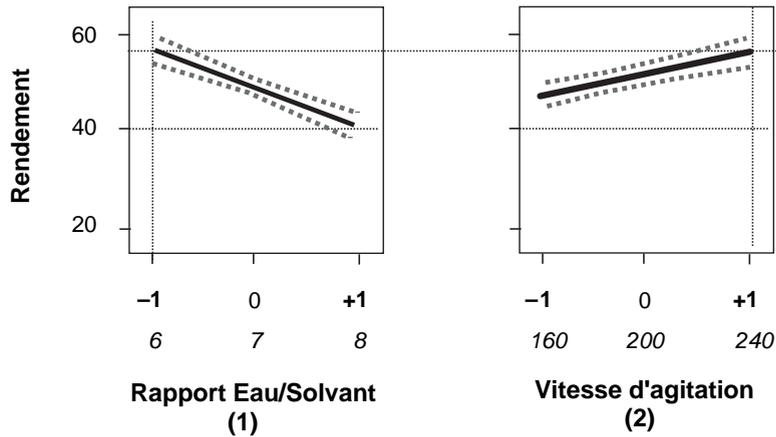


Figure 11.13 – Le meilleur rendement est obtenu quand le rapport eau/solvant est au niveau -1 (60 %) et la vitesse d'agitation au niveau +1 (240 tr/min).

11.9.4 Conclusion de l'étude

La conclusion de l'étude est très simple : on peut améliorer de manière significative le rendement de fabrication de la DHA en choisissant un rapport eau/solvant organique de 60 % et une vitesse d'agitation de 240 tr/min. L'expérimentateur mit immédiatement ces nouveaux réglages en pratique.

12 • CARRÉS LATINS ET PLANS ASSOCIÉS

Les carrés latins figurent parmi les premiers plans d'expériences introduits dans l'expérimentation. Ils sont surtout utilisés pour les variables discrètes. Le plus simple des carrés latins permet d'étudier 3 facteurs prenant 3 niveaux en 9 essais au lieu de 27. Il s'agit donc, si l'on étend la notation des plans factoriels fractionnaires aux carrés latins, d'un plan 3^{3-1} . On appelle aussi *carrés latins* des plans ayant toujours 3 facteurs mais possédant 4 et 5 niveaux par facteurs. On peut aussi imaginer des plans de 4 facteurs ayant 4 niveaux par facteurs. Il s'agit alors de plans *gréco-latins* ou plus simplement de plans 4^{4-2} qui ne comprennent que 16 essais au lieu de 256 pour le plan complet 4^4 . La dénomination *carré latin* vient du fait que les premières représentations de ces plans utilisaient des lettres latines pour désigner le troisième facteur. La dénomination *gréco-latin* vient du fait que les niveaux du troisième facteur étaient désignés par des lettres latines et ceux du quatrième facteur par des lettres grecques. Si le nombre de niveaux dépasse 5, on parle de *carrés de Youden* en hommage au chimiste qui les a introduits. On peut avoir, par exemple, un plan de 2 facteurs prenant 7 niveaux ou un plan de 3 facteurs prenant 8 niveaux. On peut poursuivre l'extension de ces plans à un nombre quelconque de facteurs et de niveaux.

Les variables discrètes doivent être modélisées et interprétées d'une manière particulière. En effet, les facteurs ne forment plus un espace continu et la notion de distance disparaît. Il convient donc avant d'aborder les carrés latins et les plans associés de définir les modèles mathématiques employés, d'introduire la notion d'effet d'un facteur et d'examiner les outils d'interprétation.

12.1 Exemple 19 : les salaires – un facteur discret à plusieurs niveaux

Nous nous appuyerons sur l'exemple 19 sur les salaires, pour commencer cette étude des plans pour facteurs discrets.

12.1.1 Modèle mathématique

Prenons l'exemple de trois personnes qui ont gagné des salaires différents. Jacques a gagné 90 €, Lee 120 € et Simon 180 €. Le salaire est modélisé par la relation :

$$y_i = \bar{y} + a_i + e_i \quad \{12.1\}$$

- y_i est le salaire de la personne i : c'est la réponse.
- \bar{y} est la moyenne des trois salaires.

- a_i est l'effet la personne i .
- e_i est un terme d'écart.

L'effet de Jacques est a_J , l'effet de Lee a_L et l'effet de Simon a_S (Figure 12.1). Les effets sont repérés par rapport à la moyenne. Il s'agit donc des réponses corrigées de la moyenne que nous avons rencontrées au chapitre 5 (analyse de la variance). Ces effets ne sont pas indépendants et il existe la relation :

$$a_J + a_L + a_S = 0$$

Quand on connaît deux effets, on peut calculer le troisième.

La relation {12.1} peut s'écrire :

$$y_i = \bar{y} + a_i + e_i$$

On en déduit que la somme des carrés des réponses corrigées de la moyenne est égale à la somme des carrés des effets augmentée de la somme des carrés des écarts (le terme $\sum a_i e_i$ est égal à 0) :

$$\sum (y_i - \bar{y})^2 = \sum a_i^2 + \sum e_i^2$$

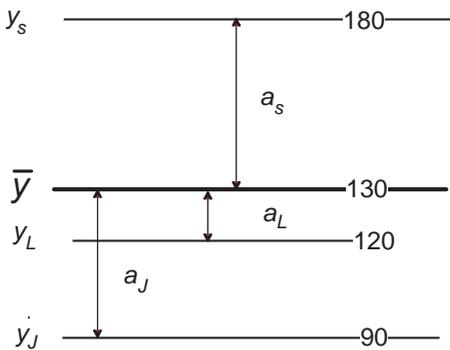


Figure 12.1 – Définition des effets des variables discrètes.

Application numérique

Reprenons l'exemple des trois personnes qui ont gagné des salaires différents. La moyenne des salaires est égale à 130 :

$$\bar{y} = \frac{1}{3}(90 + 120 + 180) = 130$$

L'effet de Jacques est : $a_J = y_J - \bar{y} = 90 - 130 = -40$

L'effet de Lee est : $a_L = y_L - \bar{y} = 120 - 130 = -10$

L'effet de Simon est : $a_S = y_S - \bar{y} = 180 - 130 = 50$

On a bien : $a_J + a_L + a_S = -40 - 10 + 50 = 0$

Dans cet exemple il y a trois équations. On peut calculer trois inconnues : la moyenne et deux effets. On peut en déduire le troisième effet, mais on ne peut pas calculer les écarts qui sont alors égaux à 0.

Application numérique

Calculons la somme des carrés des réponses corrigées de la moyenne :

$$(y_J - \bar{y})^2 + (y_L - \bar{y})^2 + (y_S - \bar{y})^2 = (-40)^2 + (-10)^2 + (50)^2 = 4\,200$$

12.1.2 Effet global d'un facteur

L'effet global des salaires ne peut pas être défini par la somme des effets qui est toujours égale à zéro. C'est pourquoi on utilise la moyenne quadratique. L'effet moyen a des salaires est défini par la relation :

$$a^2 = \frac{1}{3}[a_J^2 + a_L^2 + a_S^2]$$

D'une manière plus générale, si un facteur prend n niveaux, le carré de l'effet global est défini par :

$$a^2 = \frac{1}{n} \sum_{i=1}^{i=n} a_i^2$$

Application numérique

$$3 \sum a^2 = a_J^2 + a_L^2 + a_S^2 = (-40)^2 + (-10)^2 + (50)^2 = 4\,200$$

$$a^2 = \frac{1}{3} \times 4\,200 = 1\,400$$

$$a = 37,41$$

12.2 Suite de l'exemple 19 – deux facteurs discrets à plusieurs niveaux

12.2.1 Plan d'expériences

Poursuivons avec le même exemple sur les salaires mais supposons maintenant que la réponse dépend de deux facteurs discrets prenant chacun trois niveaux : la réponse « salaire » dépend de la qualification de la personne (Jacques a un effet a_J , Lee, un effet a_L , Simon, un effet a_S) et de l'employeur (l'entreprise A a un effet b_A , l'entreprise B, un effet b_B , l'entreprise C, un effet b_C). Jacques a travaillé chez A, puis chez B et enfin chez C. Il en est de même pour Lee et pour Simon. Les neuf points d'expériences sont disposés comme l'indique la figure 12.2. Il s'agit d'un plan 3^2 .

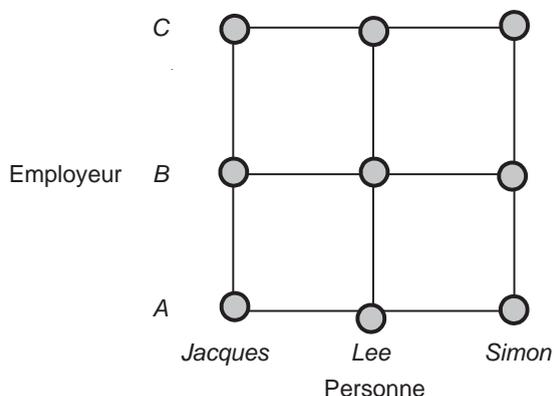


Figure 12.2 – Plan complet pour étudier deux facteurs discrets ayant chacun trois niveaux.

Les salaires gagnés par les 3 personnes dans les 3 entreprises sont consignés dans le tableau 12.1 et illustrés par la figure 12.3.

Tableau 12.1 – Matrice d'expérimentation.

Essai n°	Personne (1)	Employeur (2)	Salaire
1	Jacques	A	85
2	Jacques	B	93
3	Jacques	C	92
4	Lee	A	116
5	Lee	B	121
6	Lee	C	123
7	Simon	A	174
8	Simon	B	182
9	Simon	C	184

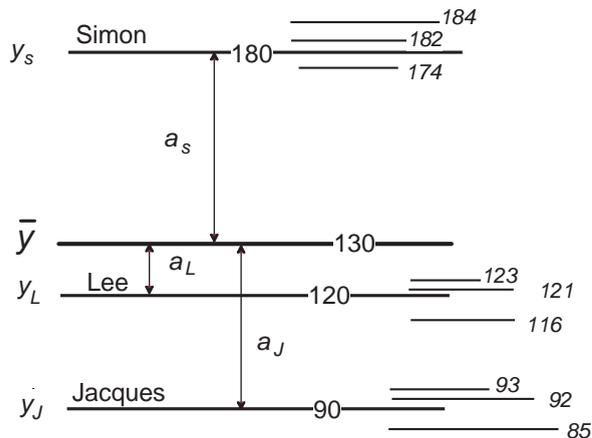


Figure 12.3 – Répartition des réponses autour des moyennes.

12.2.2 Modèle mathématique

Les salaires sont modélisés par la relation :

$$y_{i,j} = \bar{y} + a_i + b_j + e_{i,j} \quad \{12.2\}$$

– $y_{i,j}$ est le salaire de la personne i qui a travaillé dans l'entreprise j .

- \bar{y} est la moyenne des neuf salaires.
- a_i est l'effet de la personne i .
- b_j est l'effet de l'entreprise j .
- $e_{i,j}$ est un terme d'écart qui dépend de la personne et de l'entreprise.

Comme les effets sont repérés par rapport à la moyenne, ils ne sont pas indépendants et on a les relations suivantes :

$$\begin{aligned} a_J + a_L + a_S &= 0 \\ b_A + b_B + b_C &= 0 \end{aligned}$$

On introduit les gains moyens de chaque personne. Par exemple, le gain moyen \bar{y}_J de Jacques est égal à :

$$\bar{y}_J = \frac{1}{3}[y_{J,A} + y_{J,B} + y_{J,C}] \quad \{12.3\}$$

On a donc la contrainte :

$$(y_{J,A} - \bar{y}_J) + (y_{J,B} - \bar{y}_J) + (y_{J,C} - \bar{y}_J) = 0$$

On a de même les contraintes suivantes pour Lee et Simon :

$$(y_{L,A} - \bar{y}_L) + (y_{L,B} - \bar{y}_L) + (y_{L,C} - \bar{y}_L) = 0$$

$$(y_{S,A} - \bar{y}_S) + (y_{S,B} - \bar{y}_S) + (y_{S,C} - \bar{y}_S) = 0$$

Des contraintes analogues existent pour les entreprises puisqu'on introduit également les salaires moyens versés par les entreprises :

$$(y_{J,A} - \bar{y}_A) + (y_{L,A} - \bar{y}_A) + (y_{S,A} - \bar{y}_A) = 0$$

$$(y_{J,B} - \bar{y}_B) + (y_{L,B} - \bar{y}_B) + (y_{S,B} - \bar{y}_B) = 0$$

$$(y_{J,C} - \bar{y}_C) + (y_{L,C} - \bar{y}_C) + (y_{S,C} - \bar{y}_C) = 0$$

L'interprétation est basée d'une part sur les effets de chaque facteur et d'autre part sur l'analyse des carrés des réponses corrigées de la moyenne.

12.2.3 Effet des facteurs

On repère les effets à partir de la moyenne de toutes les réponses \bar{y} :

$$\bar{y} = \frac{1}{9}[85 + 93 + 92 + 116 + 121 + 123 + 174 + 182 + 184] = 130$$

L'effet de Jacques est la moyenne de ses gains \bar{y}_J par rapport à la moyenne générale :

$$a_J = \bar{y}_J - \bar{y} = \left[\frac{1}{3}(85 + 92 + 93) \right] - 130 = -40$$

Les effets de Lee et de Simon sont calculés de la même manière. On trouve $a_L = -10$ et $a_S = 50$.

L'effet de l'entreprise A est la moyenne des salaires qu'elle a distribués par rapport à la moyenne générale :

$$b_A = \bar{y}_A - \bar{y} = \left[\frac{1}{3}(85 + 116 + 174) \right] - 130 = -5$$

Les effets des entreprises B et C sont calculés de la même manière. On trouve $b_B = 2$ et $b_C = 3$.

On peut donc dresser le tableau des effets des deux facteurs étudiés (Tableau 12.2).

Tableau 12.2 – Effets des facteurs.

Facteurs	Valeurs
Moyenne	130
Jacques	-40
Lee	-10
Simon	50
Entreprise A	-5
Entreprise B	2
Entreprise C	3

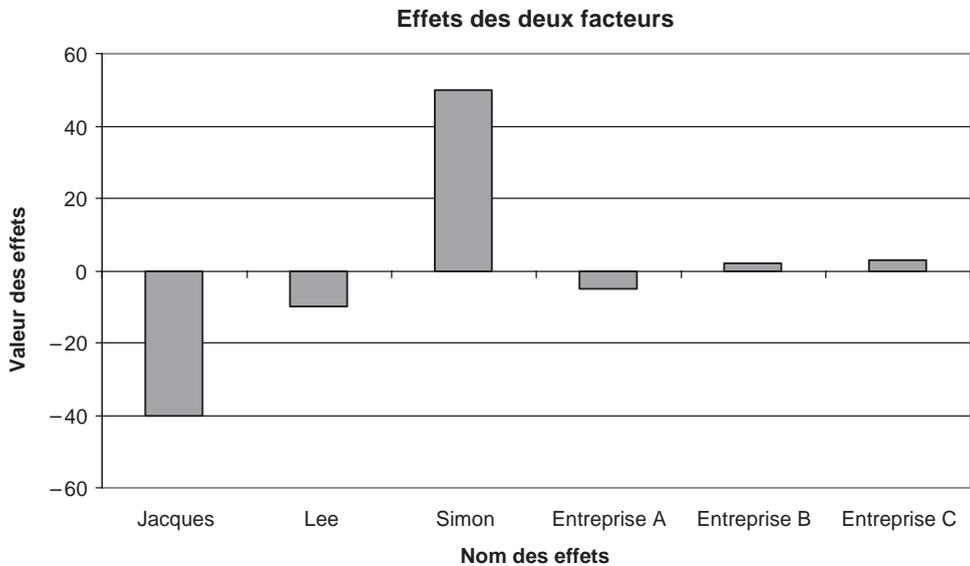


Figure 12.4 – Diagramme à barres des effets des facteurs : personnes et entreprises.

12.2.4 Analyse des sommes de carrés

Élevons les relations {12.2} au carré et additionnons ces carrés :

$$\sum (y_{i,j} - \bar{y})^2 = \sum a_i^2 + \sum b_j^2 + \sum e_{i,j}^2$$

La somme des carrés des réponses mesurées corrigées de la moyenne est égale aux sommes des carrés des effets augmentées de la somme des carrés des écarts.

Application numérique

1. Calcul de $\sum (y_{i,j} - \bar{y})^2$. Calculons les réponses mesurées corrigées de la moyenne et leurs carrés (Tableau 12.3).

Tableau 12.3 – Calcul des carrés des réponses mesurées corrigées de la moyenne.

Essai n°	Réponses initiales	Moyenne	Réponses corrigées de la moyenne	Carré des réponses corrigées de la moyenne
1	85	130	-45	2 025
2	93	130	-37	1 369
3	92	130	-38	1 444
4	116	130	-14	196
5	121	130	-9	81
6	123	130	-7	49
7	174	130	44	1 936
8	182	130	52	2 704
9	184	130	54	2 916

Puis additionnons ces carrés. On trouve que la somme des carrés des réponses mesurées corrigées de la moyenne est égale à 12 720.

2. Calcul de $\sum a_i^2$. Calculons les effets des personnes et leurs carrés pour tous les essais (Tableau 12.4).

Tableau 12.4 – Calcul des carrés des effets des personnes.

Essai n°	Moyennes des salaires touchés	Moyenne générale	Effets des individus	Carré des effets des individus
1	90	130	-40	1 600
2	90	130	-40	1 600
3	90	130	-40	1 600
4	120	130	-10	100
5	120	130	-10	100
6	120	130	-10	100
7	180	130	50	2 500
8	180	130	50	2 500
9	180	130	50	2 500

Additionnons ces carrés. On trouve que la somme des carrés des effets des personnes est égale à 12 600.

3. Calcul de $\sum b_j^2$. L'entreprise A a payé 85 € à Jacques, 166 € à Lee et 174 € à Simon, soit en moyenne 125 €. L'effet de l'entreprise A est donc de $125 - 130 = -5$ €. De même, l'effet de l'entreprise B est de $132 - 130 = 2$ € et celui de l'entreprise C de $133 - 130 = 3$ €.

Calculons les carrés de ces effets des entreprises pour tous les essais (Tableau 12.5).

Tableau 12.5 – Calcul des carrés des effets des entreprises.

Essai n°	Moyennes des salaires versés	Moyenne générale	Effets des entreprises	Carré des effets des entreprises
1	125	130	-5	25
2	132	130	2	4
3	133	130	3	9
4	125	130	-5	25
5	132	130	2	4
6	133	130	3	9
7	125	130	-5	25
8	132	130	2	4
9	133	130	3	9

Additionnons ces carrés. On trouve que la somme des carrés des effets des entreprises est égale à 114.

4. Calcul de $\sum e_{i,j}^2$. Calculons la différence entre les réponses mesurées et les réponses calculées :

$$e_{J,A} = y_{J,A} - (\bar{y} + a_J + b_A) = 85 - (130 - 40 - 5) = 0$$

$$e_{J,B} = y_{J,B} - (\bar{y} + a_J + b_B) = 93 - (130 - 40 + 2) = 1$$

$$e_{J,C} = y_{J,C} - (\bar{y} + a_J + b_C) = 92 - (130 - 40 + 3) = -1$$

$$e_{L,A} = y_{L,A} - (\bar{y} + a_L + b_A) = 116 - (130 - 10 - 5) = 1$$

$$e_{L,B} = y_{L,B} - (\bar{y} + a_L + b_B) = 121 - (130 - 10 + 2) = 1$$

$$e_{L,C} = y_{L,C} - (\bar{y} + a_L + b_C) = 123 - (130 - 10 + 3) = 0$$

$$e_{S,A} = y_{S,A} - (\bar{y} + a_S + b_A) = 174 - (130 + 50 - 5) = 1$$

$$e_{S,B} = y_{S,B} - (\bar{y} + a_S + b_B) = 182 - (130 + 50 + 2) = 0$$

$$e_{S,C} = y_{S,C} - (\bar{y} + a_S + b_C) = 184 - (130 + 50 + 3) = -1$$

D'où la somme des carrés des résidus :

$$\begin{aligned} \sum e_{i,j}^2 &= e_{JA}^2 + e_{JB}^2 + e_{JC}^2 + e_{LA}^2 + e_{LB}^2 + e_{LC}^2 + e_{SA}^2 + e_{SB}^2 + e_{SC}^2 \\ \sum e_{i,j}^2 &= 0^2 + 1^2 + (-1)^2 + 1^2 + 1^2 + 0^2 + 1^2 + 0^2 + (-1)^2 = 6 \end{aligned}$$

On rassemble tous les carrés que nous venons de calculer dans le tableau de l'analyse des sommes de carrés (Tableau 12.6). La somme des carrés des réponses mesurées est décomposée en trois sommes de carrés : celle due aux personnes, celle due aux entreprises et celle due aux résidus. Le facteur qui possède la somme de carrés la plus élevée est le facteur le plus influent. Dans le cas de cet exemple, les résultats montrent que la qualification des personnes est le facteur le plus influent sur les salaires reçus. Le modèle explique 12 714 sur les 12 720 à expliquer ($R^2 = 0,9995$). C'est un bon modèle.

Tableau 12.6 – Analyse des sommes de carrés.

Source des variations	Somme des carrés
Personnes	12 600
Entreprises	114
Résidus	6
Total (réponses mesurées)	12 720

On additionne les sommes des carrés des facteurs pour avoir la somme des carrés du modèle. Ici, la somme des carrés du modèle est égale à la somme des carrés des personnes augmentée de celle des entreprises :

$$\text{Somme des carrés du modèle} = 12\,600 + 114 = 12\,714$$

12.2.5 Analyse de la variance

Si l'on introduit les degrés de liberté, on peut transformer le tableau 12.6 en un tableau d'analyse de la variance. En effet, les variances sont les sommes des carrés des réponses corrigées de la moyenne divisée par le nombre de degrés de liberté correspondant.

Il y a 9 réponses mesurées qui sont mathématiquement indépendantes. Les réponses mesurées corrigées de la moyenne ne sont plus indépendantes puisqu'on peut calculer la dernière avec les 8 premières et la moyenne. La somme des carrés des réponses mesurées corrigées de la moyenne ne possède plus que 8 degrés de liberté. Les effets des personnes ne possèdent que 2 degrés de liberté puisqu'on les repère par rapport à leur moyenne. Il en est de même des effets des entreprises. La somme des carrés du modèle possède donc 4 degrés de liberté. En tenant compte des contraintes qui existent entre les réponses et les moyennes qui servent à calculer les résidus, on en conclut que le nombre de degrés de liberté des résidus est de 4.

Ces éléments permettent de dresser le tableau de l'analyse de la variance (Tableau 12.7).

Tableau 12.7 – Analyse de la variance.

Source des variations	Somme des carrés	ddl	Carrés moyens	F de Fisher moyens
Modèle	12 714	4	3 178,50	2 119
Résidus	6	4	1,5	Prob F < 0,0001
Total (réponses mesurées corrigées de la moyenne)	12 720	8		

Le F de Fisher est très élevé, on peut en conclure que le modèle explique bien les réponses mesurées.

12.3 Carrés latins

Si l'on veut étudier 3 facteurs prenant chacun 3 niveaux, il faudrait exécuter $3^3 = 27$ essais dans le cas d'un plan complet. Ce nombre d'essais est beaucoup trop important et les expérimentateurs ont cherché à le réduire. Des plans fractionnaires ont été proposés. Au lieu de faire 27 essais on peut construire un plan de 9 essais. On a ainsi un plan 3^{3-1} . Pour réussir la construction d'un tel plan, il ne faut pas disposer les points d'expériences au hasard. On part d'un plan 3^2 et on attribue les trois lettres latines a , b et c aux points d'expériences de telle manière que l'on retrouve a , b et c sur toutes les lignes et sur toutes les rangées sans qu'il y ait deux fois la même lettre (Figure 12.5).

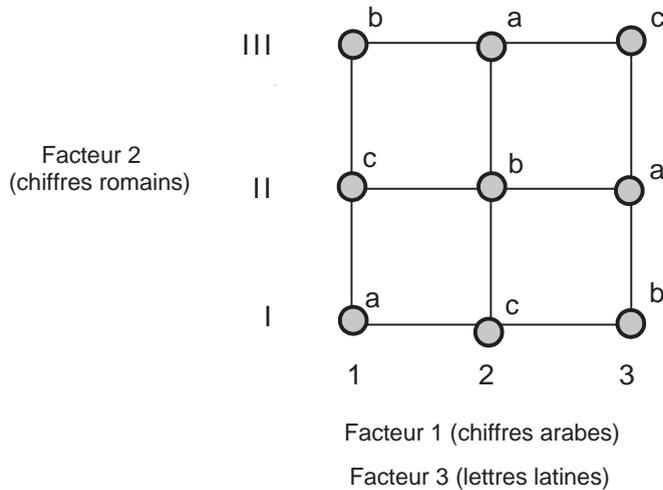


Figure 12.5 – Carré latin pour étudier trois facteurs ayant chacun trois niveaux.

Il y a douze arrangements possibles des points d'expériences qui respectent la disposition de ces carrés latins. Un moyen possible de randomiser les essais, est de choisir l'un de ces arrangements au hasard comme plan d'expériences.

On peut aussi donner une représentation tridimensionnelle de ce plan (Figure 12.6).

S'il y a 3 facteurs prenant chacun 4 niveaux, il faudrait exécuter $4^3 = 64$ essais dans le cas d'un plan complet. Le plan en carré latin propose de n'exécuter que 16 essais. Là aussi les points ne sont pas disposés au hasard (Figure 12.7).

Ces plans sont surtout utilisés pour les variables discrètes et le modèle mathématique est le suivant :

$$y_{i,j,k} = \bar{y} + a_i + b_j + c_k + e_{i,j,k}$$

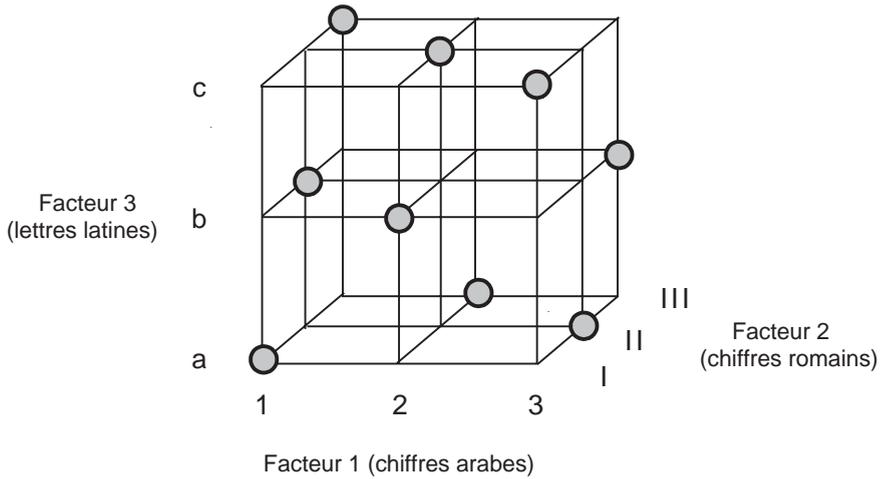


Figure 12.6 – Vue en trois dimensions d’un plan en carrés latins.

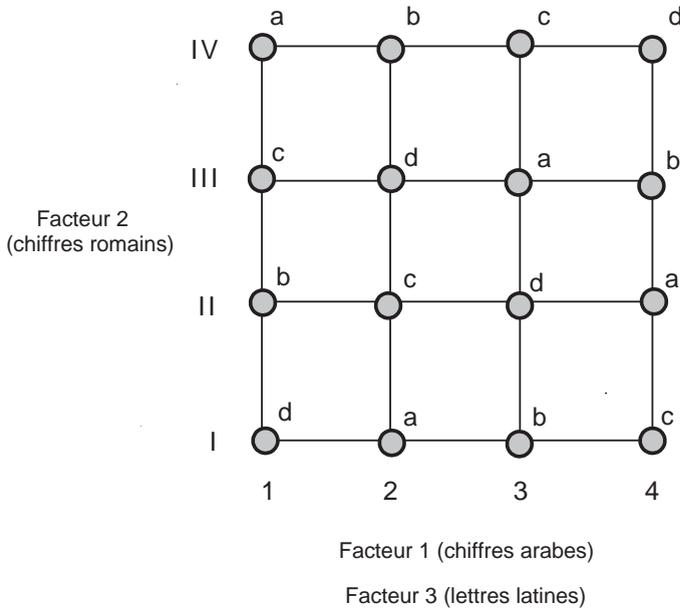


Figure 12.7 – Carré latin pour étudier 3 facteurs ayant chacun 4 niveaux.

12.4 Carrés gréco-latins

Les carrés gréco-latins sont des plans de 16 essais permettant d’étudier 4 facteurs ayant chacun 4 niveaux. Ce sont des plans 4^{4-2} (Figure 12.8). Les emplacements des points expérimentaux sont tels que chaque lettre, latine ou grecque, soit sur toutes les lignes et sur toutes les rangées sans qu’il n’y ait deux fois la même lettre.

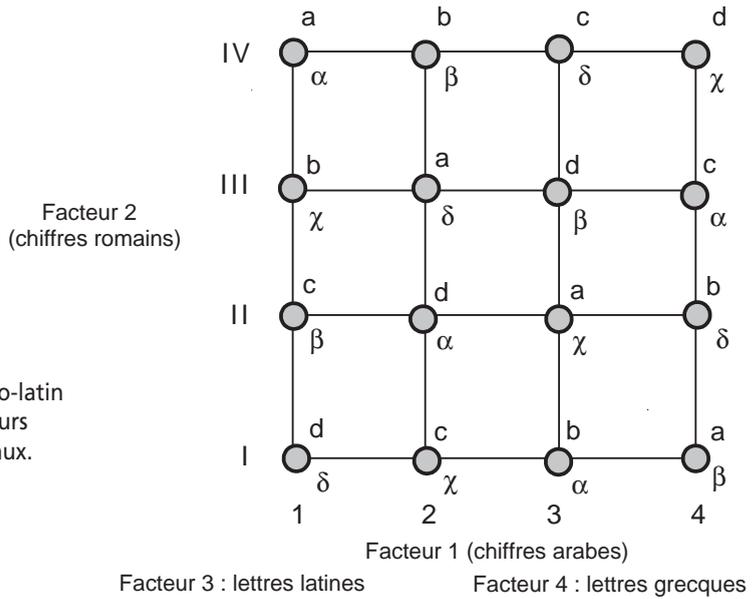


Figure 12.8 – Carré gréco-latin pour étudier 4 facteurs ayant chacun 4 niveaux.

Ces plans sont utilisés pour les variables discrètes et le modèle mathématique est :

$$y_{i,j,k,l} = \bar{y} + a_i + b_j + c_k + d_l + e_{i,j,k,l}$$

12.5 Carrés de Youden

Youden a développé des plans pour deux ou plusieurs variables discrètes prenant plus de quatre niveaux (Figure 12.9). Le principe est analogue à celui des carrés latins. On réduit le nombre des essais en retirant des points au plan complet. Les points expérimentaux sont disposés comme pour les carrés latins et les carrés gréco-latins.

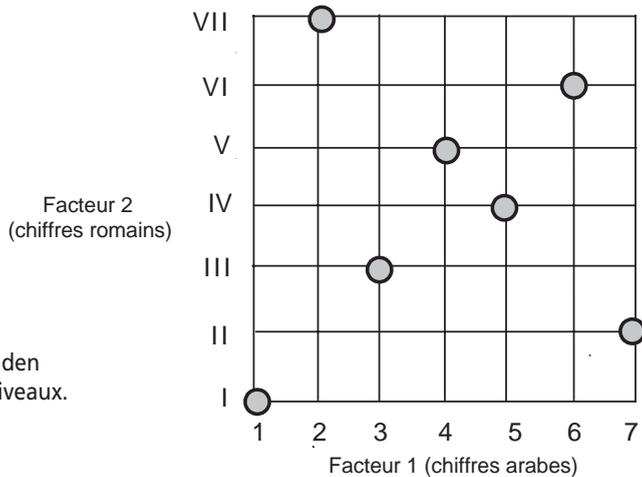


Figure 12.9 – Plan de Youden pour 2 facteurs prenant 7 niveaux.

Au lieu d'avoir un essai par niveau, on peut en avoir deux et plus (Figure 12.10).

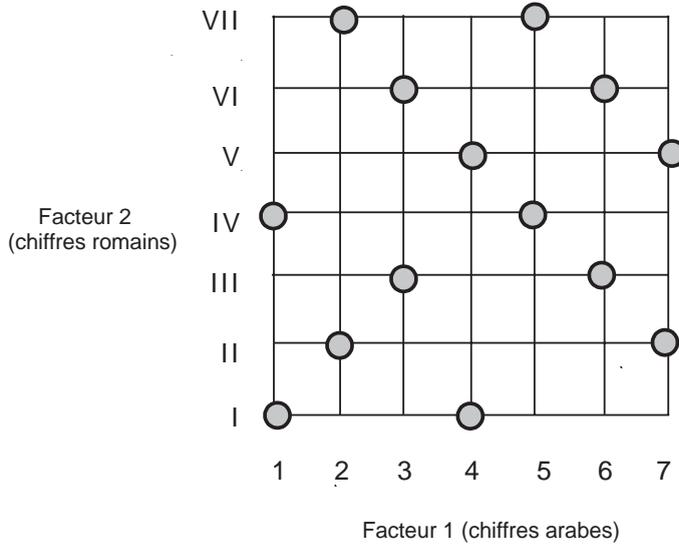


Figure 12.10 – Plan de Youden pour 2 facteurs ayant 2 essais pour chacun des 7 niveaux.

12.6 Exemple 20 : étude de la pénétrométrie

12.6.1 Préparation du plan d'expériences

Un expérimentateur désire sélectionner une méthode permettant de mesurer la consistance d'un gel. Il enfonce un stilet dans la matière à une vitesse et à une profondeur données. Il mesure la résistance au mouvement opposée par la matière gélifiée. Il étudie trois formes de stilet (cylindre, cône et disque), trois vitesses (lente, moyenne et rapide) et trois profondeurs (faible, moyenne et forte) (Tableau 12.8). Il effectue un plan en carré latin de 9 essais au lieu d'effectuer les 27 combinaisons possibles. L'objectif de l'étude est de trouver la forme de stilet et les conditions d'utilisation qui maximisent la résistance au mouvement.

Tableau 12.8 – Facteurs et domaine d'étude.

Facteur	Niveau bas	Niveau moyen	Niveau haut
Forme (1)	Cylindre	Cône	Disque
Vitesse (2)	Lente	Moyenne	Rapide
Profondeur (3)	Faible	Moyenne	Forte

12.6.2 Expérimentation

Les résultats de l'expérimentation sont donnés dans le tableau 12.9.

Tableau 12.9 – Matrice d'expériences du plan en carré latin.

Essai n°	Forme (1)	Vitesse (2)	Profondeur (3)	Résistance
1	–	–	–	17,1
2	–	0	0	20,2
3	–	+	+	27,0
4	0	–	0	8,7
5	0	0	+	10,6
6	0	+	–	7,4
7	+	–	+	62,3
8	+	0	–	70,8
9	+	+	0	86,1
–	Cylindre	Lente	Faible	
0	Cône	Moyenne	Moyenne	
+	Disque	Rapide	Forte	

12.6.3 Interprétation des résultats

L'interprétation est basée sur le calcul des effets individuels des facteurs (Tableau 12.10) et sur l'analyse de la variance (Tableau 12.11). On constate que la forme est le facteur le plus influent. Parmi les trois formes, c'est le disque qui offre la résistance la plus élevée.

Tableau 12.10 – Effet des facteurs.

Facteurs	Effet
Moyenne	34,5
Cône	–25,6
Cylindre	–13
Disque	38,6

Tableau 12.10 – Effet des facteurs. (Suite)

Facteurs	Effet
Vitesse lente	-5,1
Vitesse moyenne	-0,6
Vitesse rapide	5,7
Profondeur faible	-2,7
Profondeur moyenne	3,9
Profondeur forte	-1,2

Ce tableau peut prendre l'aspect d'un graphique (Figure 12.11).

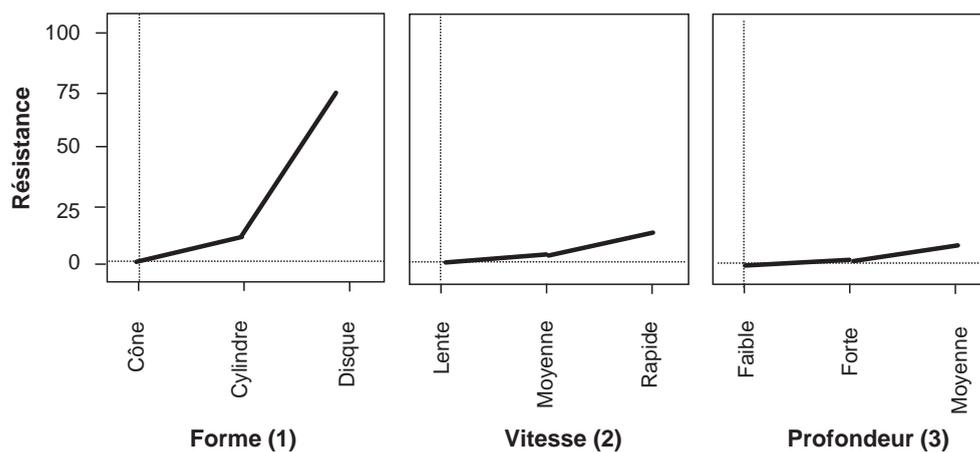


Tableau 12.11 – Analyse de la variance.

Source des variations	Somme des carrés	ddl	Carrés moyens	F de Fisher	Probabilité
Formes	6 940,4467	2	3 470,22335	69,40	0,0142
Vitesse	176,5800	2	88,29	1,76	0,3616
Profondeur	70,8067	2	35,40335	0,708	0,5855
Résidus	100,0067	2	50		
Total (réponses mesurées)	7 287,8400	8			

12.6.4 Conclusion de l'étude

On peut conclure que c'est le facteur « forme » qui contribue le plus aux variations de la réponse et c'est le disque qui oppose la plus grande résistance au déplacement du stylet. Pour faire les mesures, on choisira donc un disque. Les deux autres facteurs « vitesse » et « profondeur » ont peu d'influence. La résistance semble toutefois être augmentée par une vitesse rapide. Les essais de contrôle ont montré qu'un disque tournant rapidement permettait effectivement d'obtenir une bonne précision sur les mesures.

13 • SYNTHÈSE ET CONSEILS

La méthodologie décrite dans cet ouvrage pour conduire des recherches et organiser des essais apporte souplesse, précision, sûreté, gain de temps et économie d'argent dans toute expérimentation. On peut se demander si elle n'a que des qualités et s'il ne faut pas se poser la question de ses limites et de son domaine de validité. Dans ce chapitre nous voulons faire la synthèse de toutes les connaissances acquises : nous examinerons ce que la méthode des plans d'expériences peut nous apporter et ce qu'elle ne peut pas faire. Nous en profiterons pour donner quelques conseils et montrer les développements que l'on peut envisager pour approfondir la méthode des plans d'expériences.

La démarche utilisée pour mener à bien, et dans les meilleures conditions, toute expérimentation peut être résumée dans un schéma de synthèse (Figure 13.1), schéma que nous allons commenter et qui reprend les trois grandes parties de la méthodologie expérimentale parfois appelée *expérimentique* :

- Choix d'une méthode d'expérimentation.
- Analyse des résultats.
- Acquisition progressive des connaissances.

13.1 Choix d'une méthode d'expérimentation

Ce choix doit être effectué après une série de questions qui visent à bien poser le problème et à ne rien oublier avant de réaliser les premières expériences.

13.1.1 Bien poser le problème

Cette phase de l'étude ne doit jamais être négligée. Au contraire, il ne faut pas hésiter à y consacrer du temps et à faire appel à tous les spécialistes concernés. Toutes les connaissances du sujet traité doivent être rassemblées, analysées, critiquées et évaluées. Il arrive souvent que le problème soit modifié après une analyse en profondeur des connaissances initiales. L'essentiel est de démarrer l'étude dans les meilleures conditions possibles en intégrant tous les renseignements disponibles. C'est ainsi que l'on accélère les recherches et que l'on arrive rapidement à des conclusions justes et précises. Négliger cette phase c'est aller au devant de mauvaises surprises et se réserver de cuisants déboires.

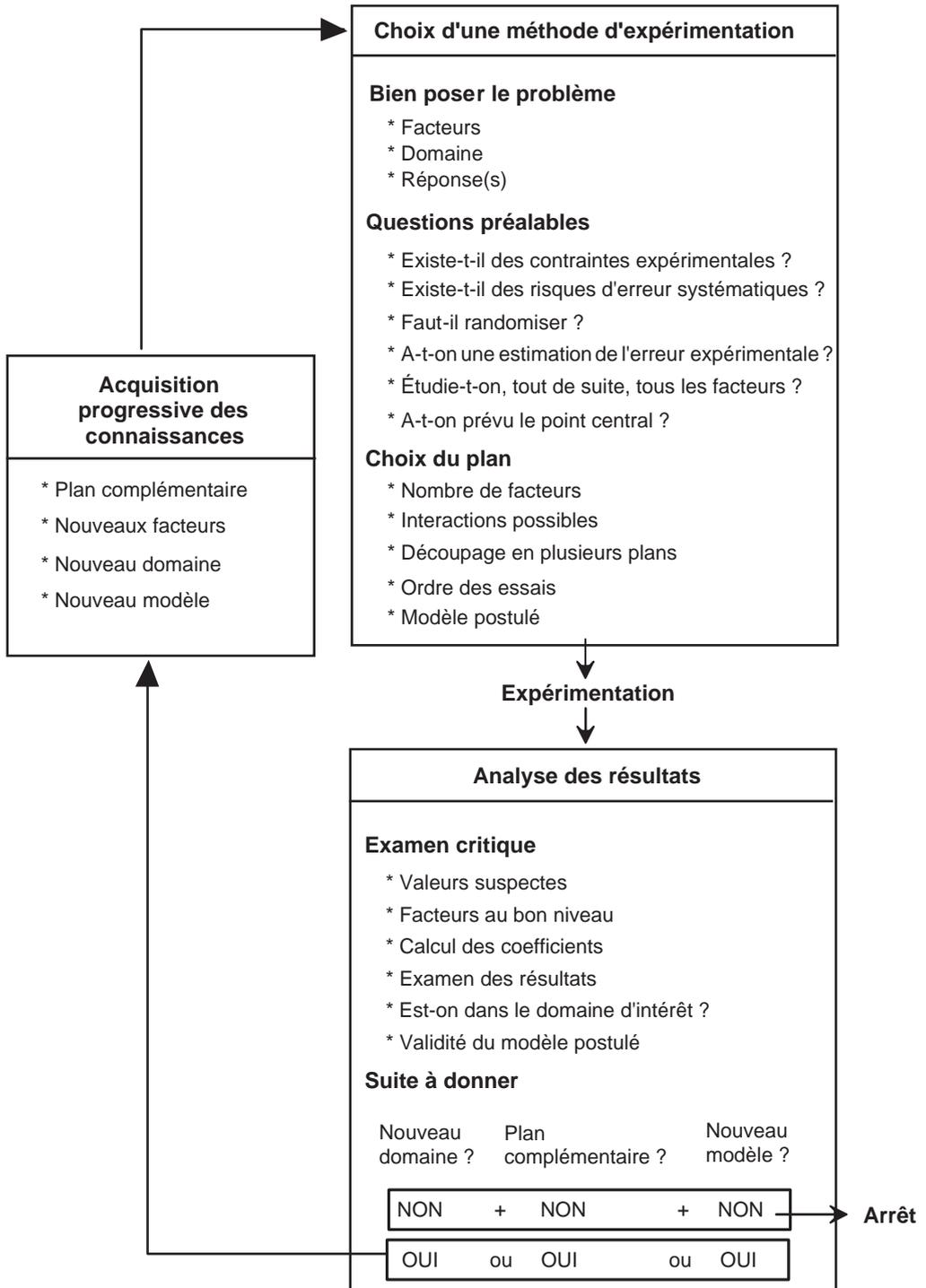


Figure 13.1 – Schéma général de conduite d'une étude.

■ Objectif de l'étude

Il est essentiel de bien préciser l'objectif de l'étude. Il y a des études générales d'acquisition de connaissances, des études particulières où il faut atteindre un but précis, des études d'optimisation d'une ou plusieurs réponses. Selon l'objectif on ne réalisera ni les mêmes essais, ni les mêmes plans d'expériences. Il faut, à chaque fois que l'on peut, définir l'objectif sous une forme quantifiable. Ne pas dire par exemple « obtenir un bon rendement » mais « obtenir un rendement au moins égal à 80 % ». À la fin de l'étude, il faut être capable de dire si l'objectif est atteint ou non. Dans ce dernier cas, il faut pouvoir dire si l'on est loin ou près du but.

■ Facteurs

Il faut lister tous les facteurs pouvant intervenir sur le phénomène étudié sans s'autocensurer. Nous avons vu que le nombre de facteurs n'est pas une limite à l'expérimentation et que les plans fractionnaires permettent de réduire le nombre des expériences. Il faut vaincre les mauvaises habitudes prises avec la méthode classique du « un facteur à la fois » et ne pas hésiter à étudier tous les facteurs pouvant influencer les réponses. Il faut abandonner les impasses et les hypothèses restrictives. Pour ne rien oublier, on pourra utiliser le moyen mnémotechnique des 6M supporté par le diagramme d'Ishikawa :

Méthodes, Machines, Main d'œuvre, Milieu, Matières, Mesures.

■ Domaine

La délimitation des niveaux de chaque facteur sera également examinée avec attention. Le domaine ne doit pas être trop petit pour que les conclusions aient une certaine généralité. Il ne doit pas être trop grand non plus, car les modèles mathématiques postulés risquent alors de ne plus être adaptés. Le meilleur compromis est difficile à trouver si l'on n'a pas suffisamment de connaissances sur le phénomène étudié. Il est conseillé, dans ce cas, de procéder à une étude préalable pour orienter le choix des niveaux haut et bas de chaque facteur.

■ Réponse(s)

Le choix de la ou des réponses est une étape capitale. Toute l'analyse et toutes les conclusions dépendent de ce choix. Une réponse mal adaptée au problème rendra l'expérimentation inutilisable ou conduira à des conclusions sans intérêt. La réponse sélectionnée doit permettre de répondre sans ambiguïté à la question posée. Ce choix n'est pas aussi simple que l'on pourrait le croire et c'est parfois la principale difficulté à laquelle on se heurte. Par exemple, nous avons passé beaucoup de temps et dépensé des trésors d'imagination pour trouver une réponse permettant de bien mesurer le pétillant d'une eau minérale ou pour évaluer le stress des contrôleurs du ciel.

13.1.2 Questions préalables

Avant de commencer les expériences, il est nécessaire de se poser un certain nombre de questions pour orienter le choix du plan initial et des éventuels plans (ou essais) complémentaires. Voici une liste, non limitative, des questions à se poser :

■ Existe-t-il des contraintes expérimentales ?

Si un réglage est difficile ou si une mise en température est longue à atteindre, il faudra envisager de conduire les essais par séquences. On effectuera, par exemple, une série d'essais au niveau bas du réglage (ou de la température) suivie d'une série d'essais au niveau haut du réglage (ou de la température). Un seul facteur pourra être étudié de cette manière. Par exemple dans le cas d'un plan 2^3 , c'est la colonne 3 qu'il faut choisir pour étudier ce facteur particulier lorsque l'ordre des essais est l'ordre habituel. Dans les matrices d'Hadamard il y a toujours une colonne, mais il n'y en a qu'une, qui est constituée d'une série de signes moins suivie d'une série de signes plus (ou inversement). La position de cette colonne dépend de l'ordre dans lequel les essais ont été choisis.

■ Existe-t-il des risques d'erreurs systématiques ?

Nous avons vu que les erreurs systématiques (chapitre 8) avaient une répercussion importante sur l'évaluation des effets et des interactions. Il est donc important de se prémunir contre ces erreurs et d'en atténuer les conséquences en choisissant convenablement l'ordre des essais.

Si l'erreur systématique risque de modifier les coefficients du modèle, il faudra opérer un blocking. C'est le cas, par exemple, d'un plan effectué à deux périodes séparées par un long intervalle de temps ou d'un plan exécuté en deux lieux différents. Le blocking permet d'obtenir les effets comme si la moyenne ne changeait pas entre les deux périodes ou les deux localisations. S'il y a risque de dérive, il faudra :

- soit choisir un ordre des essais tel que les effets principaux ne soient pas affectés d'un biais dû à cette erreur systématique ;
- soit contrôler la dérive par des mesures au point central (ou autre point de contrôle) et corriger les réponses en conséquence ; cette seconde solution est plus précise mais plus coûteuse en nombre d'expériences.

■ Faut-il randomiser ?

Le but de la randomisation est d'essayer de répartir au hasard les petites erreurs systématiques et de pouvoir ainsi traiter l'erreur expérimentale comme si elle n'avait qu'une composante : l'erreur aléatoire. Cette procédure est donc utile pour pouvoir appliquer les tests statistiques, cependant elle a l'inconvénient de gonfler l'erreur expérimentale.

La bonne démarche consiste à tenir compte de toutes les contraintes et à effectuer les arbitrages en trois temps :

- *Premier temps* : on tient compte des contraintes expérimentales : séquence de niveaux bas/séquence de niveaux haut pour un facteur. À l'intérieur de ces séquences il est encore possible de randomiser. Tous les autres arrangements permettant de simplifier les manipulations sont évidemment possibles et peuvent être envisagés par l'expérimentateur.
- *Deuxième temps* : on prend en compte les erreurs systématiques : on choisit un ordre spécial pour tenir compte d'une dérive, connue ou éventuelle, ou on découpe le plan en plusieurs blocs (blocking).

- *Troisième temps* : on randomise ce qui peut l'être encore. Par exemple :
 - on peut très bien randomiser les essais à l'intérieur de chaque bloc après avoir effectué un blocking ;
 - on peut également choisir au hasard l'un des ordres qui permet d'obtenir les effets principaux non entachés de l'erreur de dérive ;
 - s'il n'y a pas de contraintes expérimentales et s'il n'y a pas de raison de choisir un ordre particulier (dérive) ou d'effectuer un blocking, on randomisera tous les essais du plan.

■ A-t-on une estimation de l'erreur expérimentale ?

La connaissance de l'erreur expérimentale est importante pour pouvoir apprécier si un effet est significatif ou non. L'idéal est de posséder une estimation de cette erreur basée sur un grand nombre de mesures. Malheureusement, cette situation n'est pas souvent remplie et l'on est contraint de faire des hypothèses pour obtenir une évaluation approchée de cette erreur.

La comparaison effet/écart-type n'est pas aussi simple que beaucoup de manuels le laissent supposer et il ne faut pas oublier les nombreuses hypothèses, la plupart du temps cachées, pour faire cette comparaison valablement. En effet, aux hypothèses faites pour calculer une erreur expérimentale, il faut ajouter celles que l'on fait pour calculer les effets, en particulier :

- La valeur des écarts ne dépend pas de la valeur de la réponse. Cette hypothèse est rarement vérifiée.
- La distribution des écarts est une distribution normale. Il existe des tests pour vérifier si la distribution des écarts est normale.
- Les écarts ne sont pas corrélés entre eux.
- Absence d'erreurs systématiques. Cette hypothèse est rarement vérifiée.
- L'écart-type est le même dans tout le domaine d'étude (homoscédasticité). Cette hypothèse est rarement vérifiée.
- Indépendance des variables. Il peut arriver que deux variables paraissent indépendantes alors qu'en réalité il existe une relation cachée qui les lie. Cette liaison cachée peut être introduite par les propriétés physico-chimiques des composants intervenant dans l'étude.
- Le modèle mathématique postulé n'est qu'une approximation du modèle réel de la nature.
- Le modèle mathématique associé aux plans factoriels est additif.

Cette liste n'est probablement pas complète, mais elle montre la fragilité des raisonnements que l'on peut faire car, dans la plupart des études, ces hypothèses ne sont pas toutes respectées. L'expérimentateur devra donc examiner avec soin la validité des calculs d'erreur et se méfier des méthodes automatiques. Il faut bien comprendre que pour obtenir une information utile tout en diminuant le nombre des essais il a fallu faire des hypothèses simplificatrices. Ces hypothèses sont utiles et même indispensables mais, pour qui les oublie, le danger est grand de commettre une bévue ou pire une catastrophe.

Si l'on ne possède pas d'estimation de l'erreur et si l'on a la possibilité d'effectuer quelques répétitions, la meilleure solution est de réaliser des expériences au point

central ou sur tout autre point d'intérêt si le point central est inaccessible. Nous n'insisterons jamais assez sur l'utilité de ces mesures.

■ Étudie-t-on, tout de suite, tous les facteurs ?

Il se peut que la liste des facteurs à étudier soit longue et que les moyens (temps et budget) ne permettent pas de réaliser beaucoup d'essais. Dans ce cas, on commencera par sélectionner les facteurs qui risquent d'être les plus influents. Les autres facteurs momentanément non étudiés se verront attribuer un niveau fixe pendant toute l'expérimentation. Par exemple, si l'on a recensé dix facteurs et que l'on ne peut réaliser que huit essais, il faudra choisir un plan 2^{7-4} pour étudier sept facteurs et assigner un niveau constant aux trois autres pendant tout le déroulement des expériences. On se réserve ainsi la possibilité de les réintroduire ensuite dans l'étude sans perdre les huit premiers essais : on leur attribue un niveau différent de celui auquel ils avaient été fixés et l'on refait un plan analogue au premier. Cette opération est l'inverse du fractionnement : on part d'un plan 2^k et on lui ajoute 1, 2, 3... p facteurs supplémentaires, on réalise de cette manière un plan 2^{k+p} .

■ A-t-on prévu le point central ?

Le point central du domaine est un point privilégié. C'est en ce point que les réponses calculées sont, en général, obtenues avec la meilleure précision. C'est aussi le point où la réponse expérimentale est égale à la constante du modèle.

Si l'on effectue plusieurs mesures au point central, on pourra également avoir une idée de l'erreur expérimentale.

Pour ces raisons, nous conseillons fortement de réaliser quelques mesures au point central dès que cela est possible ou sur tout autre point de contrôle si le point central n'existe pas. C'est une sage précaution qui valide les hypothèses faites ou qui oriente la suite de l'expérimentation vers la bonne direction : modèle du second degré dans la majorité des cas.

13.1.3 Choix du plan

Le choix du plan sera guidé par la discussion de tous les points précédents :

- le nombre de facteurs,
- les interactions pouvant être significativement différentes de zéro,
- plan complet ou plan fractionnaire en fonction des contraintes expérimentales,
- l'ordre des essais qui sera défini,
- un point de contrôle qui sera le plus souvent prévu avec ou sans répétition.

Ce plan sera éventuellement découpé en sous-plans (blocking, plans fractionnaires) pour ne pas engager tout le budget (temps et/ou argent) dans une seule voie de recherche. On respectera ainsi le principe de l'acquisition progressive des connaissances.

Si un facteur n'est pas étudié immédiatement mais risque de l'être par la suite, il faut le maintenir à un niveau fixe. On prépare ainsi la suite de l'expérimentation. Il suffira, comme nous l'avons déjà vu, de lui donner un autre niveau et de réaliser

les nouvelles expériences correspondantes pour obtenir un plan ayant un (ou plusieurs) facteur de plus.

La mise en forme du plan initial choisi par l'expérimentateur sera facilitée par la présentation que nous avons utilisée dans tous nos exemples. Elle rassemble toutes les données : facteurs, domaine et réponses. Elle facilitera le suivi de l'expérimentation et l'analyse des résultats.

13.2 Expérimentation

C'est une phase très importante de l'étude puisque c'est la collecte des données. Les expérimentateurs apporteront donc, comme ils en ont l'habitude, le plus grand soin à réaliser les expériences. La qualité de leur travail et de leurs recommandations dépend de la qualité des données collectées. Les plans d'expériences ne les dispenseront pas de bien travailler... au contraire ! L'outil est puissant, les risques sont plus grands.

13.3 Analyse des résultats

Avant d'effectuer les calculs, l'expérimentateur examinera les conditions d'expérimentation, les résultats de ses essais et décidera d'une suite à donner.

13.3.1 Examen critique

■ Les résultats sont-ils cohérents ? N'y a-t-il pas de valeurs suspectes ?

Un résultat faux se répercutera sur tous les coefficients. Il est donc prudent de s'assurer de la qualité des résultats avant de commencer les calculs et de faire l'interprétation. Un résultat suspect peut provenir d'erreurs de transcription, d'erreurs de réglage, d'un appareil de mesure défectueux, d'un niveau mal fixé, etc.

■ Les niveaux des facteurs étaient-ils bien fixés aux valeurs prévues ?

Il peut très bien arriver que, pour un facteur, le niveau -1 (ou $+1$) soit respecté pour des essais et ne le soit pas pour d'autres. Par exemple, le niveau $+1$ du facteur 3 a été réglé à $+1,2$ au cours des essais n° 2 et 5. Les expériences ne sont pas perdues car il suffit d'entrer les bons niveaux dans la matrice d'expériences du tableur de l'ordinateur pour que les résultats soient corrects.

Lorsque l'examen critique des résultats d'expériences est terminé et que l'on est sûr que tout est correct, on procède aux calculs des coefficients et des principales statistiques courantes (écarts-types, p -values et R^2). On pourra ensuite construire les graphiques et analyser les résultats en profondeur.

■ Calcul des coefficients

Cette analyse se fait en plusieurs étapes. Ne jamais oublier que les plans fractionnaires donnent des contrastes et non des effets. C'est par commodité de langage que l'on emploie souvent le mot *effet* à la place de *contraste*. Le calcul préalable et l'examen des alias éviteront les conclusions hâtives. L'analyse est une phase longue

et délicate au cours de laquelle l'expérimentateur se livre à un examen approfondi des résultats :

- Quels sont les facteurs influents ?
- Quelles sont les interactions significatives ?
- Existe-t-il des ambiguïtés dans les contrastes calculés ?
- Le modèle mathématique associé aux plans peut-il orienter les futurs travaux ?
- La comparaison « constante du modèle/réponse mesurée au centre » valide-t-elle ou non le modèle PDAI ?
- Est-il réaliste d'avoir supposé les effets additifs ?
- Le domaine contient-il les solutions du problème posé ?
- Une analyse des résidus a-t-elle été pratiquée pour s'assurer que toute l'information restante a été extraite ?

L'ensemble de ces questions, augmenté de celles qui pourraient se poser en fonction de chaque étude, conduira l'expérimentateur à réaliser des calculs, des graphiques, à revenir sur les résultats, à effectuer des comparaisons, c'est-à-dire à faire un examen critique et en profondeur de toute son expérimentation. Le phénomène étudié aura déjà livré une grande partie de ses secrets et peut-être même suffisamment pour que l'étude soit terminée après un premier plan. Mais il arrive que des expériences supplémentaires soient nécessaires pour avancer dans la connaissance du problème posé. Plusieurs voies sont alors envisageables :

13.3.2 Suite à donner

En fonction des conclusions du premier plan, l'expérimentateur pourra se poser les questions suivantes :

- Faut-il envisager un changement de domaine ?
- Faut-il envisager un plan complémentaire (ou des essais complémentaires) ?
- Faut-il envisager un autre modèle mathématique ?

■ Domaine d'intérêt

Une extension du domaine d'étude est à envisager si les premiers résultats montrent que la solution espérée n'est pas dans le domaine défini au départ. Les résultats acquis orienteront les futurs essais vers le nouveau domaine qui devrait contenir la solution espérée.

■ Plans et essais complémentaires

Ce type de plan est envisagé lorsqu'il y a des effets principaux aliasés avec des interactions que l'on ne peut pas négliger. Il faut aussi envisager un plan complémentaire quand le modèle PDAI (modèle du premier degré avec interactions) postulé n'explique pas bien les résultats du plan. On passera alors à un modèle du second degré.

■ Le modèle postulé est-il valide ?

S'il y a de très fortes interactions ou si la constante du modèle diffère de la réponse mesurée au centre, le modèle mathématique postulé n'est pas valide. Il faut alors étayer l'interprétation sur un modèle mathématique du second degré. Si cela ne suffit pas, on construira un plan optimal adapté au modèle mathématique qui représente le mieux le phénomène étudié.

■ Faut-il arrêter les essais ?

En fonction des réponses aux questions précédentes, l'expérimentateur pourra soit arrêter les essais : il en sait assez et peut faire des recommandations répondant à l'objectif de l'étude, soit désirer des informations supplémentaires : il abordera alors le processus d'acquisition progressif des connaissances.

D'après la figure 13.1, ces situations se traduisent par :

- Il faut trois NON pour arrêter.
- S'il y a un seul OUI il faut parcourir à nouveau la boucle. Celle-ci peut, bien sûr, être parcourue plusieurs fois.

13.4 Acquisition progressive des connaissances

Si l'expérimentateur désire des informations supplémentaires, il devra définir les outils et les méthodes à utiliser. C'est en fonction de l'analyse des résultats et des questions posées qu'il choisira sa stratégie :

- *Plans ou essais complémentaires* : ces essais sont nécessaires pour désaliéser les effets principaux de certaines interactions. Ils sont aussi indispensables pour passer d'un modèle PDAI à un modèle du second degré (points en étoile). Il faut aussi envisager des essais supplémentaires pour effectuer une extension du domaine d'étude.
- *Nouveaux facteurs* : si l'on a pris la précaution de maintenir les facteurs « en attente d'examen » à un niveau fixe, on pourra les faire participer à un nouveau plan en leur assignant de nouveaux niveaux.
- *Optimisation* : le domaine d'étude peut très bien convenir mais la modélisation peut très bien être insuffisante pour traiter le problème posé. Dans ce cas, il faudra enrichir les premières expériences en définissant de nouveaux essais. Les emplacements des nouveaux points expérimentaux dans le domaine expérimental devront être choisis avec soin.

La grande souplesse des plans d'expériences permet, sans jamais perdre aucun essai :

- d'enrichir constamment par de nouvelles expériences les résultats déjà trouvés ; les plans ou les essais complémentaires affinent l'information apportée par les premiers plans ;
- d'ajouter de nouveaux facteurs si on a eu la sagesse de les prévoir et de les maintenir à un niveau fixe au cours du premier plan.

Il résulte des remarques précédentes que l'on a souvent intérêt à commencer par un plan ayant peu d'essais. Cela permettra d'obtenir rapidement de précieux renseignements pour la suite de l'étude. Puis, en fonction des objectifs visés, on effectuera les essais qui conviennent le mieux : expériences complémentaires ou facteurs supplémentaires. Le principe d'acquisition progressif des connaissances prend sa source dans ces remarques et complète harmonieusement les deux premières parties de la méthodologie de l'expérimentation : choix de la méthode d'expérimentation et analyse des résultats.

13.5 Recommandations

Les conclusions d'une étude aboutissent à des recommandations dont l'expérimentateur est le seul responsable. Il doit donc s'entourer de toutes les précautions possibles pour être sûr de ses conclusions. Le réglage des facteurs pour obtenir ou s'approcher au mieux des objectifs de l'étude est d'abord déterminé par simple calcul. Il est donc sujet à caution. Il faut absolument lever le doute et réaliser les expériences de contrôle. Ces essais de validation sont indispensables et font partie de la bonne pratique des plans d'expériences.

13.6 Limites de l'expérimentique

L'expérimentique est une démarche méthodique pour conduire avec succès une expérimentation précise en optimisant les moyens à la disposition du chercheur : temps et argent. Mais cette méthode ne fera pas le travail de l'expérimentateur et surtout ne remplacera pas son intelligence et son génie. C'est lui qui devra :

- **Définir l'objectif** : il est essentiel de bien préciser et de quantifier l'objectif de l'étude.
- **Bien poser le problème** : un problème mal posé ne sera pas redressé par l'expérimentique.
- **Être imaginatif et créatif** : l'expérimentique n'est qu'une technique, elle aide à l'interprétation des résultats mais c'est l'expérimentateur qui est responsable de ses hypothèses et de ses conclusions.
- **Choisir les facteurs** : si l'expérimentateur oublie un facteur important dans son étude, c'est lui qui sera responsable des anomalies et des conclusions erronées.
- **Définir les limites du domaine** : c'est une des grandes difficultés rencontrées par l'expérimentateur. Si le domaine est trop petit, les réponses espérées risquent de ne pas y être, ce qui entraînera des essais supplémentaires. Si le domaine est trop grand, le modèle postulé sera mal adapté pour interpréter les résultats expérimentaux.
- **Bien conduire ses expériences** : des expériences mal conçues, des facteurs non contrôlés, des mesures imprécises, des erreurs de notation, de repérage ou de transcription ne seront pas redressés par les plans d'expériences. Au contraire, ils exigent un grand soin dans l'expérimentation elle-même pour en tirer toute l'essence.
- **Choisir les réponses** : c'est également un point capital pour la qualité de l'interprétation. Toutes les réponses ne sont pas équivalentes. Certaines ne sont pas assez précises, d'autres ne fournissent pas des effets suffisamment nets pour tirer des conclusions. Enfin, il faut que les réponses choisies permettent d'atteindre l'objectif de l'étude.

En résumé, l'expérimentique est une technique utile et puissante, mais

**La maîtrise de l'expérimentation et la responsabilité des conclusions
restent entre les mains de l'expérimentateur.**

14 • PRÉPAREZ ET ANALYSEZ VOS PLANS D'EXPÉRIENCES AVEC JMP®

JMP® est un logiciel de statistique élaboré par la société SAS Institute Inc. qui est particulièrement bien adapté pour la construction et l'analyse des plans d'expériences. Il est conçu pour les ordinateurs personnels et possède des versions adaptées à Windows, Macintosh ou Linux. Il utilise une interface « pointer-cliquer », et son ensemble très complet d'outils statistiques permet la mise en forme, l'exploration et l'analyse des données expérimentales. Le présent chapitre est une courte introduction à JMP que l'on pourra compléter de la manière suivante :

- Allez sur le blog du livre : <http://www.plansdexpériences.com>. Vous y trouverez les exemples du livre accompagnés d'explications supplémentaires et des fichiers de données téléchargeables. Vous pourrez aussi télécharger un manuel d'introduction à JMP, en français, d'environ 80 pages.
- Si vous n'avez pas récupéré la documentation (version .pdf) au moment de l'installation de JMP, allez sur le site <http://www.jmp.com> sur lequel vous pourrez la télécharger. En particulier, les guides *Design of Experiments* et *Statistics and Graphics Guide* fournissent de nombreux détails sur les ressources et l'utilisation de JMP.

Nous allons décrire les tâches principales que vous pourrez effectuer avec JMP, comme construire un plan d'expériences, choisir un modèle mathématique, interpréter les résultats expérimentaux, récupérer les sorties de JMP pour les utiliser dans d'autres logiciels ou pour les incorporer dans des documents.

14.1 Construire un plan d'expériences

Dans certaines situations, les plans sur mesure sont les mieux appropriés à l'étude entreprise. Le *Custom Designer* génère alors un plan optimal adapté à votre problème et à vos moyens comme les plans étudiés au cours de cet ouvrage. Cependant, dans de nombreuses autres situations, les plans classiques (comme la plupart de ceux décrits dans ce livre) sont suffisants. Dans ce cas, on utilise d'autres moyens de construction de plans (*Screening Designer*, *Full factorial Designer*, *Response Surface Designer* et *Mixture Designer*). Nous allons voir successivement comment construire un plan avec le *Screening Designer* et le *Custom Designer*. C'est l'exemple n° 1, « Je maîtrise la consommation de ma voiture », qui est utilisé, exemple où la réponse est la consommation de carburant et où les facteurs sont la vitesse et la surcharge.

14.1.1 Screening Designer (plan de criblage classique)

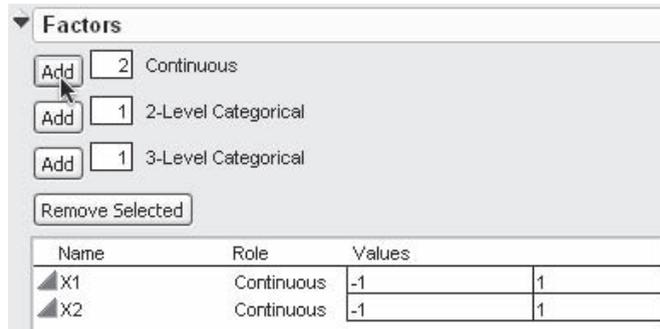
■ Entrer les réponses

Sélectionnez **DOE > Screening Design**. Lorsque la boîte de dialogue apparaît, double-cliquez sur le nom de la réponse, Y, et entrez **Consommation**.



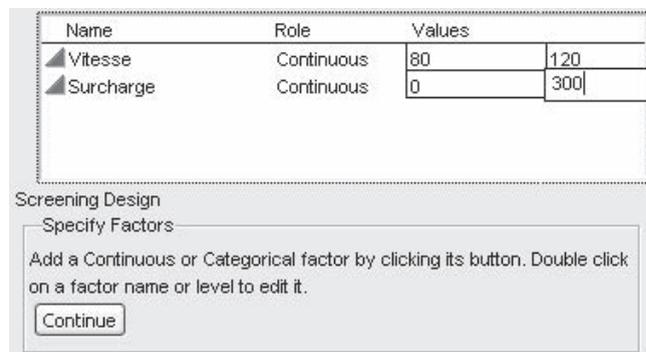
■ Entrer les facteurs

Dans la rubrique **Factors**, entrez le chiffre 2 devant **Continuous** et cliquez sur le bouton **Add** pour introduire deux facteurs.



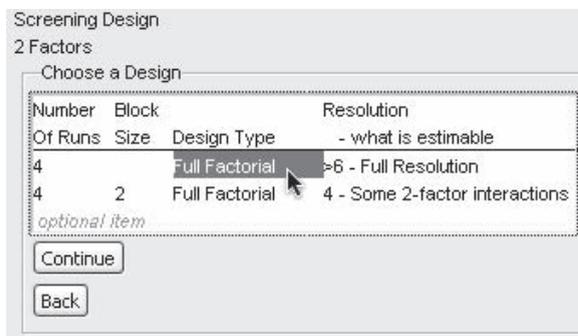
■ Modifier le nom des facteurs

Double-cliquez sur X1 et modifiez-le en **Vitesse**. Entrez les niveaux bas (**80**) et haut (**120**) de la **Vitesse**. Répétez ces opérations pour la **Surcharge**. Cliquez sur le bouton **Continue** pour passer à la rubrique suivante.



■ Choisir le plan

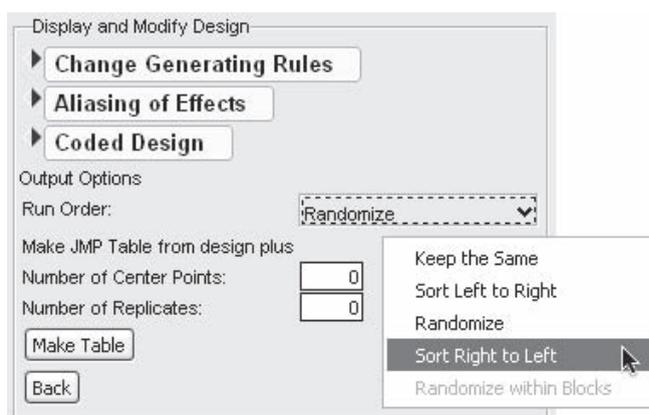
JMP vous propose plusieurs plans classiques parmi lesquels vous pouvez choisir celui qui vous convient le mieux. Vous pouvez aussi choisir d'exécuter le plan en le subdivisant en blocs. Ici, choisissez le premier plan de la liste. Il possède 4 essais et n'a aucun bloc. Cliquez sur le bouton **Continuer**.



■ Choisir les aliasés et l'ordre des essais

Vous voyez maintenant le panneau de contrôle qui vous permet de modifier votre plan. Par exemple, pour changer les aliasés, ouvrez la rubrique **Change Generating Rules** et utilisez les cases pour définir vos propres aliasés.

Vous devez maintenant sélectionner l'ordre des essais. **Randomize** est le choix par défaut. Ici, choisissez **Sort Right to Left** pour obtenir l'ordre de Yates.



■ Construire le plan

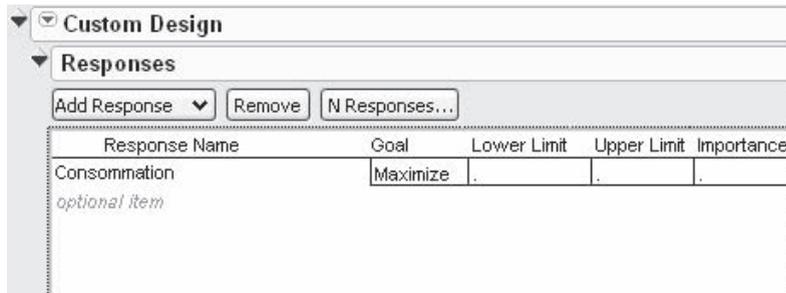
Cliquez sur le bouton **Make Table** pour générer votre plan sous la forme d'un tableau : les facteurs sont en colonnes et les essais en lignes. La colonne des réponses sera remplie après l'exécution des expériences pour pouvoir effectuer l'analyse des résultats.

Design	Full Factorial	Model	Pattern	Vitesse	Surcharge	Consommation
1	--		80	0		▪
2	+-		120	0		▪
3	--+		80	300		▪
4	++		120	300		▪

14.1.2 Custom Designer

■ Entrer les réponses

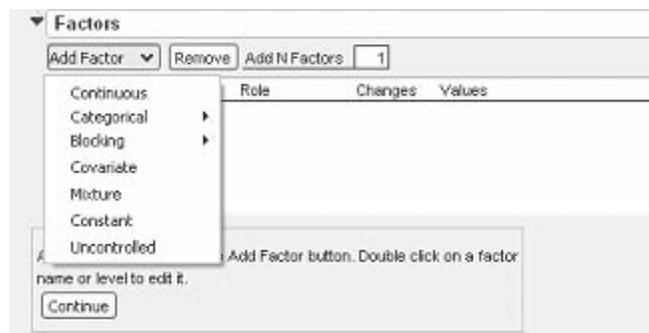
Sélectionnez **DOE > Custom Design**. Lorsque la boîte de dialogue apparaît, double-cliquez sur le nom de la réponse et écrivez **Consommation**.



■ Entrer les facteurs

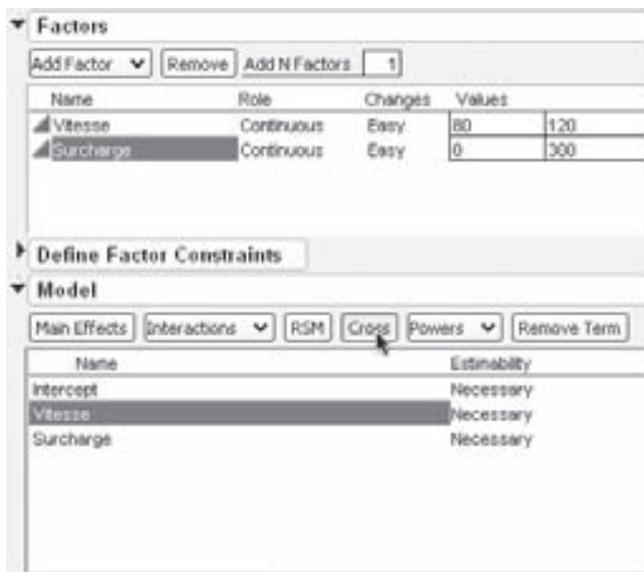
Puis cliquez sur le bouton **Add Factor** pour introduire un facteur. Choisissez **Continuous**. Indiquez le nom du facteur et ses niveaux. Répétez l'opération pour introduire le second facteur.

Il faut ensuite préciser le modèle mathématique. JMP propose un modèle avec la constante et les facteurs principaux. Vous pouvez, si besoin est, ajouter les interactions et les puissances.



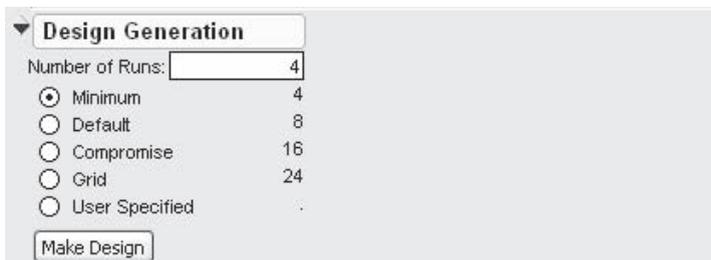
■ Indiquer les interactions

Pour entrer une interaction, surlignez le premier facteur concerné dans la rubrique **Factors**, et le second dans la rubrique **Model**, puis cliquez sur **Cross**. Si vous désirez simplement toutes les interactions d'un certain ordre, sélectionnez cet ordre grâce au bouton **Interactions** sans rien sélectionner d'autre. JMP écrira toutes les interactions.



■ Choisir le nombre d'essais

Le modèle étant spécifié, JMP vous propose plusieurs choix pour le nombre d'essais. Vous choisissez celui qui vous convient grâce aux boutons radio ou vous imposez le vôtre grâce à la case **Number of Runs**. Ici, vous mettez 4, le nombre d'essais de l'exemple. Cliquez sur **Make Design** pour générer le plan.



■ Propriétés du plan

Maintenant vous pouvez voir le plan et quatre rubriques qui contiennent des informations sur la qualité du plan. Deux de ces rubriques, **Prediction Variance Profile** et **Relative Variance of Coefficients**, sont illustrées ici.

Vous sélectionnez ensuite l'ordre des essais. **Randomize** est le choix par défaut. Ici, choisissez **Sort Right to Left** pour obtenir l'ordre de Yates. Cliquez sur le bouton **Make Table** pour générer votre plan sous forme de tableur.

The screenshot shows the JMP Design window. The 'Design' section displays a table with 4 runs. The 'Output Options' section has 'Run Order' set to 'Randomize'. A dropdown menu is open, showing 'Sort Right to Left' selected. The 'Make Table' button is visible. To the right, there are two plots: 'Prediction Variance Profile' and 'Prediction Variance Surface', and a table of 'Effect Variance'.

Run	Vitesse	Surcharge
1	120	300
2	120	0
3	80	0
4	80	300

Effect	Variance
Intercept	0.25
Vitesse	0.25
Surcharge	0.25
Vitesse*Surcharge	0.25

■ Plan

Le plan est affiché dans un tableur. Les facteurs sont en colonnes et les essais en lignes. La colonne des réponses sera remplie après l'exécution des expériences pour pouvoir effectuer l'analyse des résultats. Il n'y a pas de colonne **Pattern**.

Custom Design			Vitesse	Surcharge	Consommation
Design	Custom Design				
Criterion	D Optimal				
Model					
Columns (3/0)					
	Vitesse *	1	80	0	▪
	Surcharge *	2	120	0	▪
	Consommation *	3	80	300	▪
		4	120	300	▪

14.2 Le tableur de JMP

Lorsque vous cliquez sur **Make Table** vous obtenez un tableur (Figure 14.1). Il est divisé en deux parties : à gauche, il y a une zone d'informations, à droite, la grille des données.

Zone de dé-sélection

Nom des facteurs

Variables discrètes

Variables continues

Variable ordonnables

Numéro des essais

	popcorn	oil amt	batch	yield	trial
1	plain	little	large	8.2	1
2	gourmet	little	large	8.6	1
3	plain	lots	large	10.4	1
4	gourmet	lots	large	9.2	1
5	plain	little	small	9.9	1
6	gourmet	little	small	12.1	1
7	plain	lots	small	10.6	1
8	gourmet	lots	small	18.0	1
9	plain	little	large	8.8	2
10	gourmet	little	large	8.2	2
11	plain	lots	large	8.8	2
12	gourmet	lots	large	9.8	2
13	plain	little	small	10.1	2
14	gourmet	little	small	15.9	2
15	plain	lots	small	7.4	2
16	gourmet	lots	small	16.0	2

Figure 14.1 – Exemple de plan affiché dans le tableur de JMP.

14.2.1 Zone d'information

La zone d'informations est divisée en plusieurs sections :

- La section située en haut, appelée **Tables**, indique le nom du fichier (ici *Popcorn*) ; elle contient un ou plusieurs scripts (ici, un seul script appelé *Fit Model*) et peut fournir d'autres informations sur le tableur.
- La deuxième section, appelée **Columns**, indique les colonnes du tableur et le type de chaque variable. Le type des variables est très important car il est pris en compte par le logiciel pour effectuer les bonnes analyses et les bons graphiques. Il est possible de changer le type des variables en cliquant sur l'icône qui est située à gauche de l'intitulé de la colonne et en sélectionnant le type qui convient dans le menu qui apparaît. Trois options sont disponibles pour le type des variables : *Continuous* (désigne une variable continue par un triangle bleu), *ordinal* (facteur discret ordonnable illustré par un diagramme à barres vert dont les marches sont croissantes) et *nominal* (facteur discret illustré par un diagramme à barres rouge dont les marches sont irrégulières). Lorsque les variables sont de type *Continuous*, JMP prend la précaution de bien présenter les graphiques et les statistiques des facteurs continus. Il en est de même pour les variables *ordinal* et *nominal* qui gouvernent les graphiques et les statistiques correspondants.
- La zone suivante, appelée **Rows**, donne des informations sur les lignes du tableur. Il est important de se souvenir, surtout lorsqu'on entre les données à la main, que chaque ligne du tableur représente un essai expérimental. Il en résulte que la réponse introduite par l'expérimentateur doit parfaitement correspondre à l'essai concerné.

14.2.2 Grille des données

Il s'agit du plan proprement dit représenté sous forme d'un tableau. Les facteurs sont disposés en colonnes et les essais expérimentaux en lignes.

Les lignes et les colonnes peuvent être sélectionnées avec la souris. Pour désélectionner les lignes ou les colonnes, il faut cliquer dans la zone de désélection qui se trouve à gauche des titres de colonnes (Figure 14.1). Si l'on clique au-dessus de la diagonale, on désélectionne les colonnes et, si l'on clique au-dessous de la diagonale, on désélectionne les lignes. On peut aussi ajouter ou supprimer des lignes ou des colonnes.

14.3 Choisir et utiliser un modèle

Si vous avez construit votre plan à la main ou si vous voulez rester maître de votre analyse, vous devez remplir vous-même la fenêtre de modélisation (Figure 14.2).

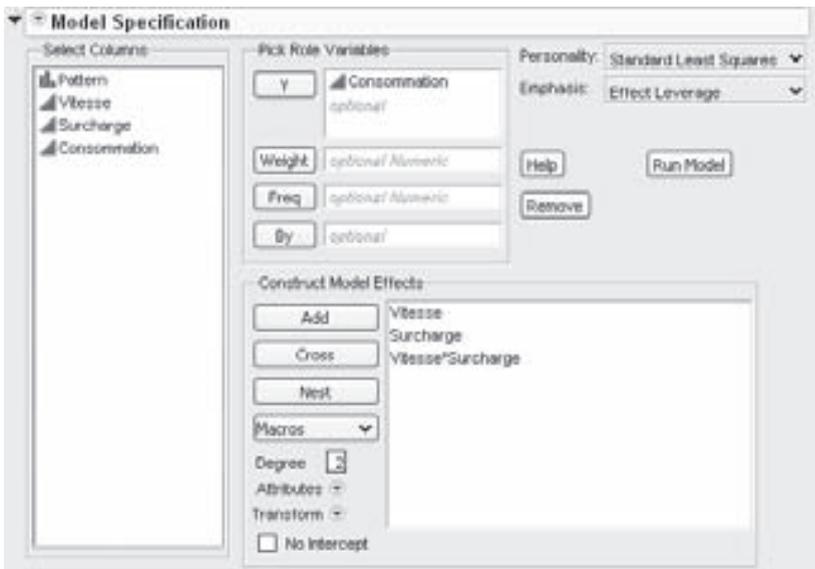


Figure 14.2 – Fenêtre de modélisation de JMP.

- Surlignez la (ou les) réponse(s) dans la plage **Select Columns** et cliquez sur le bouton **Y** qui se trouve dans la plage **Pick Role Variables**. Vous pouvez aussi utiliser le « tirer-déposer » en prenant le nom d'une réponse et en le déposant dans la plage voulue.
- Surlignez le (ou les) facteur(s) dans la plage **Select Columns** et cliquez sur le bouton **Add** qui se trouve dans la plage **Construct Model Effects**.
- Si vous voulez ajouter des interactions au modèle, surlignez les facteurs impliqués dans l'interaction dans **Select Columns** et cliquez sur le bouton **Cross**. Pour sélectionner plusieurs facteurs en même temps, maintenez appuyée la touche

Control (Windows et Linux) ou ⌘ (Macintosh) puis cliquez sur le nom des facteurs.

- Après avoir introduit les facteurs, les réponses, défini le modèle et choisi l'analyse statistique, cliquez sur le bouton **Run Model** pour lancer les calculs de l'analyse statistique.

Voici quelques conseils supplémentaires pour utiliser au mieux la boîte de dialogue *Fit Model* :

- Pour enlever un facteur du modèle, double-cliquez sur ce facteur. Vous pouvez aussi sélectionner le facteur et cliquer sur le bouton **Remove**.
- Contrairement à beaucoup de boîtes de dialogue de JMP, la boîte *Fit Model* ne disparaît pas après avoir cliqué sur **Run Model**. Cela vous permet d'examiner successivement plusieurs modèles. Les résultats d'un premier modèle guident votre analyse vers un autre modèle peut-être meilleur. Par exemple, si vous constatez qu'un modèle de plan complet possède des termes d'interaction qui sont négligeables, vous pouvez les retirer du modèle et relancer l'analyse en utilisant le bouton **Run Model**.
- Certains modèles sont fréquemment utilisés. Par exemple, si vous avez trois facteurs A, B et C, et si vous désirez faire l'analyse avec le modèle ayant toutes les interactions, il serait pénible d'entrer toutes les interactions car il faudrait sélectionner beaucoup de colonnes et effectuer de nombreux clics. Ces manœuvres sont simplifiées par l'utilisation du menu **Macros** qui propose les modèles les plus courants. Par exemple, pour entrer un modèle polynomial complet pour trois facteurs, commencez par sélectionner les trois facteurs dans la plage **Select Columns** et cliquez sur **Macros > Full Factorial** ou sur **Macros > Factorial Sorted**. Vous obtiendrez directement un modèle comprenant les facteurs principaux, toutes les interactions d'ordre deux et l'interaction d'ordre trois. De même, **Factorial to degree** fournit un modèle avec tous les facteurs et toutes les puissances jusqu'à celle spécifiée dans le champ **Degree** situé sous le menu **Macros**. Par défaut le degré est 2.
- Le menu **Attributes** permet de définir la nature des facteurs. Dans le livre, nous avons rencontré deux facteurs de natures différentes : les facteurs de mélanges (**Mixture**) (la somme des concentrations est égale à l'unité) et les facteurs pour surface de réponse (**Response Surface**) dont les valeurs sont utilisées pour découvrir des maxima et/ou des minima de la réponse. Le moyen le plus simple de construire les modèles correspondants est d'employer le menu **Macros**. Si vous avez des facteurs qui doivent être analysés comme ceux des modèles pour surface de réponse, sélectionnez-les et cliquez sur **Response Surface** du menu **Macros**. Tous les termes principaux, toutes les interactions d'ordre 2 et tous les termes de puissance seront automatiquement ajoutés. Cela est également vrai pour les facteurs de mélanges.
- Par défaut, JMP centre tous les facteurs. Ce centrage est utile pour certaines analyses statistiques. Cependant, dans le domaine des plans d'expériences, il est préférable d'employer des données non centrées. Pour lancer une analyse avec des variables non centrées, cliquez sur le triangle rouge situé à gauche de **Model**

Specification et décochez **Center Polynomials**. Pour être sûr de toujours obtenir des analyses avec des facteurs non centrés, à partir du menu principal, suivez le chemin **File/Preferences/Platforms/Model Dialog** et, dans les options, décochez **Center Polynomials [default on]**.

En haut et à droite de la boîte de dialogue *Fit Model*, il y a deux listes déroulantes dénommées **Personality** et **Emphasis**. Si vous obtenez un rapport d'analyse qui ne vous convient pas lorsque vous cliquez sur le bouton **Run Model**, vérifiez les indications de ces deux listes :

- Les indications de **Personality** indiquent les méthodes d'analyses statistiques qui vont être effectuées par JMP. Il apparaît, par défaut, des sélections basées sur le type des facteurs. Dans ce livre, la plupart des exemples utilisent des variables continues et l'analyse correspondante est la régression basée sur le critère des moindres carrés, soit une personnalité **Standard Least Squares**. D'autres options comme **Stepwise** ou **Manova** sont également prévues pour des variables continues et peuvent être choisies à l'aide de cette liste déroulante. Les analyses **Nominal Logistic** (régression logistique pour variables discrètes) et **Ordinal Logistic** (régression logistique pour variables ordinales) sont des choix par défaut d'autres types de variables. Elles ne sont pas abordées ici.
- Les indications de **Emphasis** indiquent les résultats d'analyses qui seront présentés. Le choix effectué ne modifie pas les calculs mais la présentation et le type de sorties. Il y a trois genres de présentation : **Effect Leverage** qui fournit les graphiques et les rapports concernant les coefficients et l'influence de chaque facteur sur la réponse, **Effect Screening** qui est conçu pour les plans ayant beaucoup de facteurs et pour lesquels l'intérêt réside dans la découverte des facteurs influents, **Minimal Report** qui sert quand on souhaite n'obtenir que des résultats numériques sans graphiques. Dans toutes les situations, il est possible de demander à JMP des analyses supplémentaires quelle que soit la personnalité choisie.

14.4 Examiner et analyser les résultats

Dès que vous avez cliqué sur **Run Model**, JMP affiche un rapport d'analyse dont la présentation est analogue à celle de la figure 14.3.

Plusieurs rubriques sont importantes à consulter. La rubrique **Analysis of Variance** (analyse de la variance) affiche le rapport F de Fisher et la p -value associée (notée **Prob > F**) pour l'ensemble du modèle. Une petite p -value indique un modèle ayant au moins un facteur significatif. La rubrique **Parameter Estimates** (estimation des coefficients) indique les valeurs des coefficients du modèle. La rubrique **Effect Tests** indique le rapport F de Fisher et la p -value pour chacun des facteurs du modèle.

À gauche du titre de chaque rubrique, il y a un petit triangle bleu qui permet, d'un clic, d'ouvrir ou de fermer la rubrique.

En plus de ce triangle bleu, signalons le triangle rouge à gauche du titre **Response Y** en haut du rapport d'analyse. Ces triangles rouges sont importants car ils ouvrent des menus donnant accès à des analyses statistiques supplémentaires et à de nombreux graphiques. On peut ainsi obtenir, parmi de nombreuses autres informations, la valeur

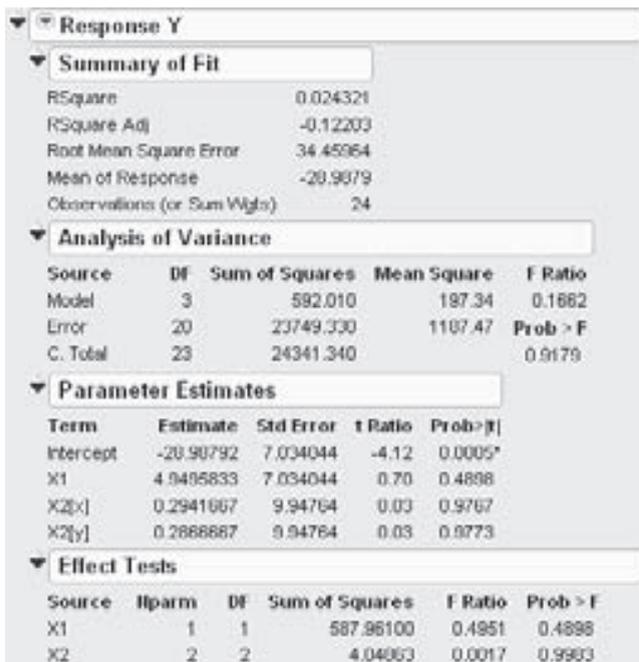


Figure 14.3 – Exemple de rapport d'analyse statistique.

des coefficients, les résidus, des graphiques sur les effets des facteurs, les corrélations entre les niveaux des facteurs, les coefficients ou les réponses, etc.

14.5 Principaux graphiques et rapports d'analyse

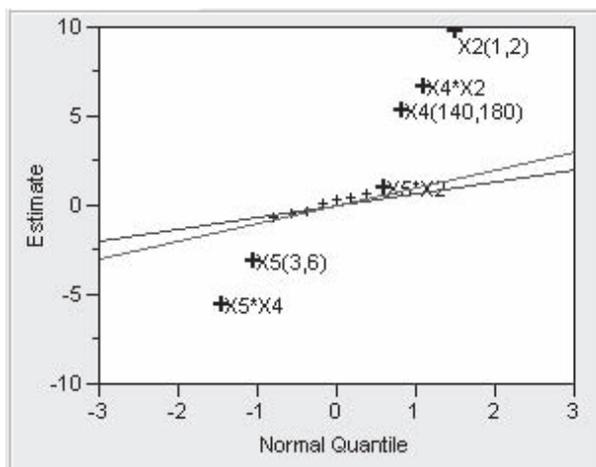
Pour les plans d'expériences les renseignements les plus importants se trouvent dans les sous-menus **Effect Screening** ou **Factor Profiling**. On pourra y consulter de nombreuses informations dont les suivantes.

Le **Normal Plot** est le diagramme de Daniel (Figure 14.4). Il indique les coefficients les plus significatifs. On peut aussi obtenir le **Half-normal Plot** qui est une présentation particulière du diagramme de Daniel : au lieu des valeurs algébriques des coefficients, il utilise les valeurs absolues.

Le **Profiler** est un graphique interactif permettant de voir l'influence des variations des niveaux des facteurs sur les réponses (Figure 14.5). Cliquez (ou tirez) les lignes rouges verticales pour voir comment les variations de niveaux des facteurs modifient la réponse.

Le triangle rouge, à gauche de la rubrique **Prediction Profiler**, donne accès à un menu qui contient la fonction désirabilité (**Desirability Functions**) vous permettant d'optimiser, grâce à la commande **Maximize Desirability**, une ou plusieurs réponses (voir chapitre 4).

La commande **Contour Profiler** fait apparaître un graphique interactif vous permettant d'étudier les courbes d'isoreponses (courbes de niveaux) d'une surface



Blue line is Lenth's PSE, from the estimates population.
Red line is RMSE, Root Mean Squared Error from the residual.

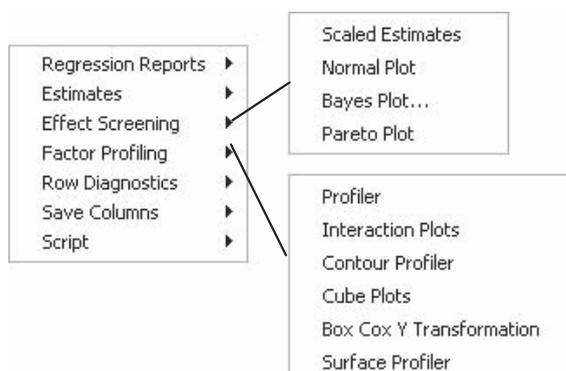


Figure 14.4 – Diagramme de Daniel et menus déroulants associés au triangle rouge du titre de la fenêtre de rapport d'analyse.

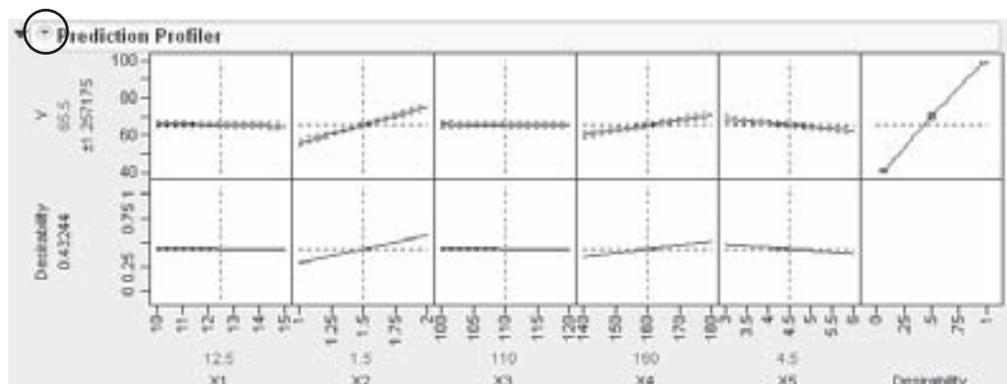


Figure 14.5 – Illustration du Prediction Profiler.

de réponse (Figure 14.6). On peut aussi accéder, avec le **Surface Profiler**, à un graphique permettant d'étudier des surfaces de réponse en trois dimensions.

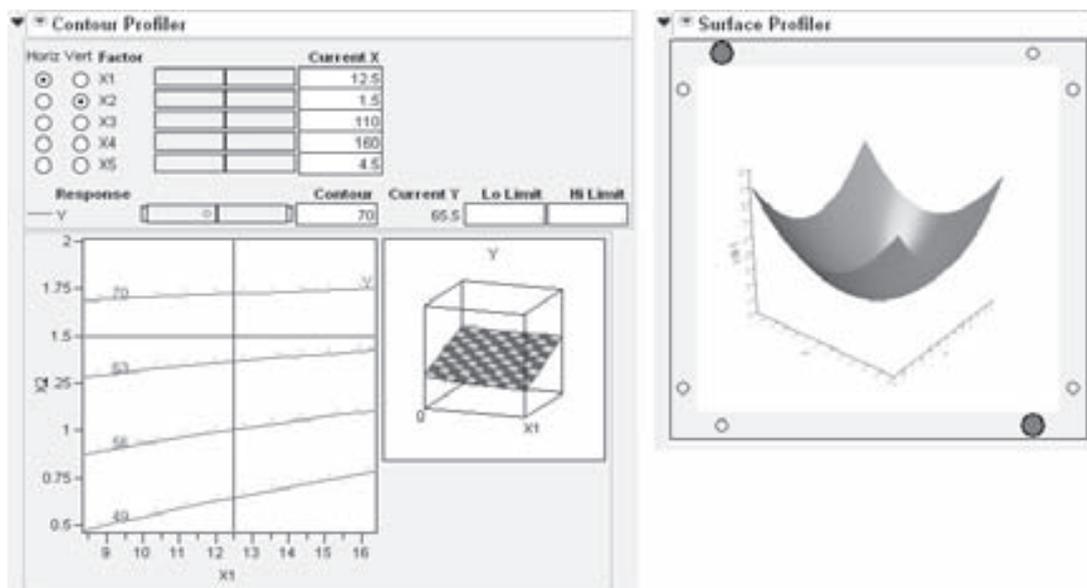


Figure 14.6 – Illustration des graphiques **Contour Profiler** (à gauche) et **Surface Profiler** (à droite).

14.6 Transférer et sauvegarder les résultats et les analyses

À la fin d'une analyse statistique, il peut être nécessaire de transférer les résultats et les graphiques de JMP vers d'autres applications, telles qu'un traitement de texte ou un logiciel de présentation. On peut aussi vouloir envoyer ces renseignements sur un site web pour une large diffusion. Il existe deux principaux moyens de transférer ces résultats.

Vous pouvez d'abord couper et coller les rapports de JMP dans une autre application en utilisant l'outil de sélection . Cet outil se trouve dans la barre d'outils ou dans le menu **Tools**. Cliquez sur l'outil et utilisez-le pour surligner la partie du rapport d'analyse qui vous intéresse. S'il y a plusieurs parties à transférer, opérez de même mais en maintenant enfoncée la touche *Control* (PC ou Linux) ou la touche  (Macintosh) pour sélectionner les différentes parties qui vous intéressent. Puis, à l'aide des commandes *Copier* et *Coller* de vos applications informatiques (que l'on trouvera la plupart du temps dans le menu *Édition*), effectuez le transfert. Les textes sont transférés comme textes et les graphiques comme graphiques.

Si vous devez transférer un rapport d'analyse ou le convertir au format HTML, effectuez les démarches suivantes :

- Sélectionnez **Edit > Journal**. Vous obtenez ainsi une copie du rapport d'analyse dans le journal de JMP. Ce journal est utile pour concaténer plusieurs rapports

différents dans une seule fenêtre. La copie ainsi effectuée peut être sauvegardée sous plusieurs formats.

- Le journal étant ouvert, sélectionnez **File > Save As** ou **File > Export** (Macintosh). Vous obtenez une fenêtre standard de sauvegarde dans laquelle plusieurs formats vous sont proposés (Figure 14.7). Utilisez le format **RTF** si vous destinez le rapport à un éditeur de texte et **HTML** si vous voulez publier une page sur le web et dont les graphiques sont stockés dans un dossier séparé.

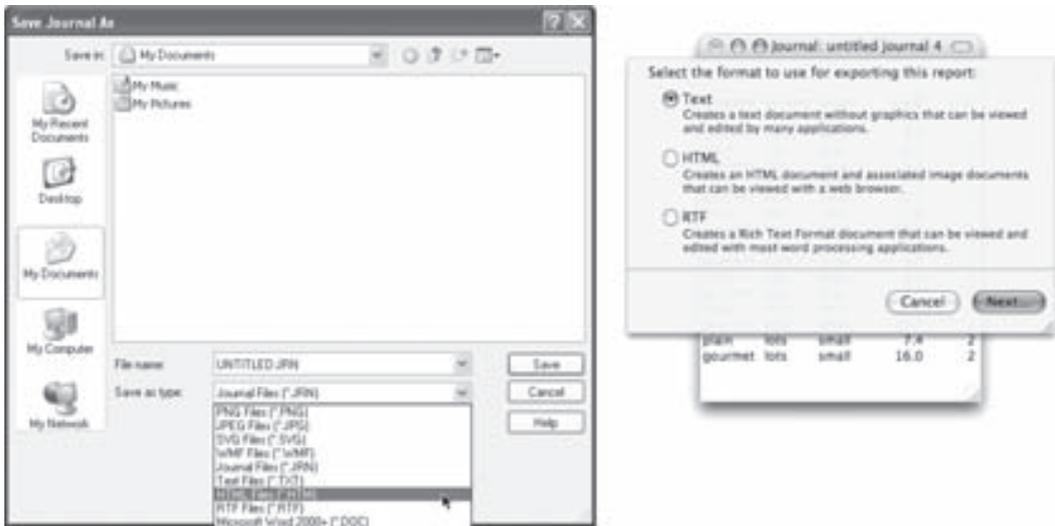


Figure 14.7 – Fenêtre de sauvegarde du journal.

Annexes

A • ORIGINE DES EXEMPLES

Exemple n°	Chap.	Titre	Origine
1	2	Je maîtrise la consommation de ma voiture	J. Goupy
2	3	Des bijoux en or	L. Chalumeau et coll., Application of a Doehlert design to the optimization of an Au-Co plating electrolyte, <i>J. of Applied Electrochemistry</i> , vol. 34, 2004, p. 1177-1184.
3	4	La galette des rois	Étude proposée par une société industrielle. Les données ont été légèrement modifiées pour assurer la confidentialité du procédé de fabrication.
4	5	Durée de vie des outils de coupe jetables	Données aimablement communiquées la société MINITAB. Les unités ne sont pas indiquées pour ne pas divulguer des renseignements confidentiels.
5	6	Le tellure	M. Grotti, E. Magi et R. Leardi, Study of interferences in graphite furnace atomic absorption spectrometry by means of experimental design, <i>Analytica Chimica Acta</i> , 327, 1996, p. 47-51.
6	7	La sulfonation	Étude extraite de la thèse de chimie réalisée par Madame Kamoun-Messedi en 2000 à l'Université de Sfax (Tunisie).
7	7	Le spectrofluorimètre	G. Sado et J. Goupy, La méthodologie des plans d'expériences appliquée à l'optimisation du réglage d'un spectrofluorimètre, <i>Analisis</i> , vol. 14, n° 8, 1986, p. 389-400.
8	7	Les pommes de terre chips	Données aimablement communiquées par la société MINITAB.

Exemple n°	Chap.	Titre	Origine
9	8	Le <i>Penicillium chrysogenum</i>	1. O.L. Davies, <i>The Design and Analysis of Industrial Experiments</i> , Oliver and Boyd, Édimbourg, 1971. 2. C. Daniel, <i>Applications of Statistics to Industrial Expérimentation</i> , John Wiley, New York, 1976.
10	8	Les haricots de Yates	F. Yates, <i>The Design and Analysis of Factorial Experiments</i> , Bulletin 35, Hafner (MacMillan), Imperial Bureau of Soil Science, Harpenden, Herts, 1937.
11	8	Le broyeur	J. Goupy, <i>Introduction aux plans d'expériences</i> , Dunod, 2 ^e édition, 2001.
12	9	La rectification du contremaître	J. Goupy, <i>Plans d'expériences pour surfaces de réponse</i> , Dunod, 1999. Plan aimablement communiqué par le Professeur P. Lantéri.
13	9	Un yoghourt doux	J. Goupy, <i>Plans d'expériences pour surfaces de réponse</i> , Dunod, 1999.
14	9	L'insecticide	J. Goupy, <i>Plans d'expériences pour surfaces de réponse</i> , Dunod, 1999. Plan aimablement communiqué par le Docteur Tralongo.
15	10	Les trois polymères	J. Goupy, <i>Plans d'expériences pour surfaces de réponse</i> , Dunod, 1999.
16	11	Développement d'un détecteur de fissures	T.E. Capobianco, J.D. Splett et H.K. Iyer, Eddy Current Probe Sensitivity as a Function of Coil Construction Parameters, <i>Research in Nondestructive Evaluation</i> , vol. 2, 1990, p. 169-186.
17	11	Les comprimés du pharmacien	J. Goupy, <i>Plans d'expériences : les mélanges</i> , Dunod, 2000.
18	11	La crème à bronzer	A. Delacroix, C. Porte et D. Youssef, <i>Application d'une méthode chimiométrique dans l'étude de la cristallisation de la dihydrocétone (DHA)</i> , congrès Chimiométrie, Paris, 1999.
19	12	Les salaires	J. Goupy
20	12	Étude de la pénétrométrie	J. Goupy

B • COMPARAISON DE DEUX MOYENNES INDÉPENDANTES

On veut comparer les moyennes de deux séries de mesures.

■ Première moyenne

On effectue n_1 répétitions au même point expérimental. On obtient une première série de mesures $y_{1,i}$, dont la moyenne \bar{y}_1 est telle que :

$$\bar{y}_1 = \frac{1}{n_1} \sum_{i=1}^{i=n_1} y_{1,i}$$

La variance des mesures $y_{1,i}$ est :

$$V(y_{1,i}) = \frac{1}{n_1 - 1} \sum_{i=1}^{i=n_1} (y_{1,i} - \bar{y}_1)^2$$

■ Deuxième moyenne

On effectue n_2 répétitions au même point expérimental. On obtient une deuxième série de mesures $y_{2,j}$, dont la moyenne est \bar{y}_2 . La variance des mesures $y_{2,j}$ est telle que :

$$V(y_{2,j}) = \frac{1}{n_2 - 1} \sum_{j=1}^{j=n_2} (y_{2,j} - \bar{y}_2)^2$$

Pour comparer \bar{y}_1 et \bar{y}_2 , on calcule :

– la valeur absolue de la différence :

$$\Delta = |\bar{y}_1 - \bar{y}_2|$$

– la variance de la différence :

$$V(\Delta) = V(\bar{y}_2) + V(\bar{y}_1)$$

– l'écart-type de la différence, soit la racine carrée de la variance :

$$\sigma_{\Delta} = \sqrt{V(\Delta)} = \sqrt{V(\bar{y}_2) + V(\bar{y}_1)}$$

– le rapport de la différence à l'écart-type de la différence :

$$\frac{\Delta}{\sigma_{\Delta}} = \frac{|\bar{y}_1 - \bar{y}_2|}{\sqrt{V(\bar{y}_1 - \bar{y}_2)}}$$

Ce rapport suit une loi de Student à $v = (n_1 - 1) + (n_2 - 1)$ degrés de liberté.

■ Hypothèses

1. Les deux séries sont issues d'une même population. Elles ont donc théoriquement la même moyenne et le même écart-type.
2. Les réponses sont issues d'une population ayant une distribution normale.

C • SÉRIES D'ESSAIS LAISSANT LES FACTEURS PRINCIPAUX INSENSIBLES À UNE DÉRIVE LINÉAIRE

Série n°	Ordre des 8 essais de chaque série							
1	1	4	6	7	8	5	3	2
2	1	4	7	6	8	5	2	3
3	1	4	8	5	6	7	3	2
4	1	4	8	5	7	6	2	3
5	1	6	4	7	8	3	5	2
6	1	6	7	4	8	3	2	5
7	1	6	8	3	4	7	5	2
8	1	6	8	3	7	4	2	5
9	1	7	4	6	8	2	5	3
10	1	7	6	4	8	2	3	5
11	1	7	8	2	4	6	5	3
12	1	7	8	2	6	4	3	5
13	1	8	4	5	6	3	7	2
14	1	8	4	5	7	2	6	3
15	1	8	6	3	4	5	7	2
16	1	8	6	3	7	2	4	5
17	1	8	7	2	4	5	6	3
18	1	8	7	2	6	3	4	5
19	2	3	5	8	7	6	4	1
20	2	3	7	6	5	8	4	1
21	2	3	7	6	8	5	1	4
22	2	3	8	5	7	6	1	4

Série n°	Ordre des 8 essais de chaque série							
23	2	5	3	8	7	4	6	1
24	2	5	7	4	3	8	6	1
25	2	5	7	4	8	3	1	6
26	2	5	8	3	7	4	1	6
27	2	7	3	6	5	4	8	1
28	2	7	3	6	8	1	5	4
29	2	7	5	4	3	6	8	1
30	2	7	5	4	8	1	3	6
31	2	7	8	1	3	6	5	4
32	2	7	8	1	5	4	3	6
33	2	8	3	5	7	1	6	4
34	2	8	5	3	7	1	4	6
35	2	8	7	1	3	5	6	4
36	2	8	7	1	5	3	4	6
37	3	2	5	8	6	7	4	1
38	3	2	6	7	5	8	4	1
39	3	2	6	7	8	5	1	4
40	3	2	8	5	6	7	1	4
41	3	5	2	8	6	4	7	1
42	3	5	6	4	2	8	7	1
43	3	5	6	4	8	2	1	7
44	3	5	8	2	6	4	1	7
45	3	6	2	7	5	4	8	1
46	3	6	2	7	8	1	5	4
47	3	6	5	4	2	7	8	1
48	3	6	5	4	8	1	2	7
49	3	6	8	1	2	7	5	4
50	3	6	8	1	5	4	2	7
51	3	8	2	5	6	1	7	4
52	3	8	5	2	6	1	4	7
53	3	8	6	1	2	5	7	4
54	3	8	6	1	5	2	4	7

Série n°	Ordre des 8 essais de chaque série							
55	4	1	5	8	6	7	3	2
56	4	1	5	8	7	6	2	3
57	4	1	6	7	5	8	3	2
58	4	1	7	6	5	8	2	3
59	4	5	1	8	6	3	7	2
60	4	5	1	8	7	2	6	3
61	4	5	6	3	1	8	7	2
62	4	5	6	3	7	2	1	8
63	4	5	7	2	1	8	6	3
64	4	5	7	2	6	3	1	8
65	4	6	1	7	5	3	8	2
66	4	6	5	3	1	7	8	2
67	4	6	5	3	7	1	2	8
68	4	6	7	1	5	3	2	8
69	4	7	1	6	5	2	8	3
70	4	7	5	2	1	6	8	3
71	4	7	5	2	6	1	3	8
72	4	7	6	1	5	2	3	8
73	5	2	3	8	4	7	6	1
74	5	2	4	7	3	8	6	1
75	5	2	4	7	8	3	1	6
76	5	2	8	3	4	7	1	6
77	5	3	2	8	4	6	7	1
78	5	3	4	6	2	8	7	1
79	5	3	4	6	8	2	1	7
80	5	3	8	2	4	6	1	7
81	5	4	2	7	3	6	8	1
82	5	4	2	7	8	1	3	6
83	5	4	3	6	2	7	8	1
84	5	4	3	6	8	1	2	7
85	5	4	8	1	2	7	3	6
86	5	4	8	1	3	6	2	7

Série n°	Ordre des 8 essais de chaque série							
87	5	8	2	3	4	1	7	6
88	5	8	3	2	4	1	6	7
89	5	8	4	1	2	3	7	6
90	5	8	4	1	3	2	6	7
91	6	1	3	8	4	7	5	2
92	6	1	3	8	7	4	2	5
93	6	1	4	7	3	8	5	2
94	6	1	7	4	3	8	2	5
95	6	3	1	8	4	5	7	2
96	6	3	1	8	7	2	4	5
97	6	3	4	5	1	8	7	2
98	6	3	4	5	7	2	1	8
99	6	3	7	2	1	8	4	5
100	6	3	7	2	4	5	1	8
101	6	4	1	7	3	5	8	2
102	6	4	3	5	1	7	8	2
103	6	4	3	5	7	1	2	8
104	6	4	7	1	3	5	2	8
105	6	7	1	4	3	2	8	5
106	6	7	3	2	1	4	8	5
107	6	7	3	2	4	1	5	8
108	6	7	4	1	3	2	5	8
109	7	1	2	8	4	6	5	3
110	7	1	2	8	6	4	3	5
111	7	1	4	6	2	8	5	3
112	7	1	6	4	2	8	3	5
113	7	2	1	8	4	5	6	3
114	7	2	1	8	6	3	4	5
115	7	2	4	5	1	8	6	3
116	7	2	4	5	6	3	1	8
117	7	2	6	3	1	8	4	5
118	7	2	6	3	4	5	1	8

Série n°	Ordre des 8 essais de chaque série							
119	7	4	1	6	2	5	8	3
120	7	4	2	5	1	6	8	3
121	7	4	2	5	6	1	3	8
122	7	4	6	1	2	5	3	8
123	7	6	1	4	2	3	8	5
124	7	6	2	3	1	4	8	5
125	7	6	2	3	4	1	5	8
126	7	6	4	1	2	3	5	8
127	8	1	2	7	3	6	5	4
128	8	1	2	7	5	4	3	6
129	8	1	3	6	2	7	5	4
130	8	1	3	6	5	4	2	7
131	8	1	5	4	2	7	3	6
132	8	1	5	4	3	6	2	7
133	8	2	1	7	3	5	6	4
134	8	2	1	7	5	3	4	6
135	8	2	3	5	1	7	6	4
136	8	2	5	3	1	7	4	6
137	8	3	1	6	2	5	7	4
138	8	3	1	6	5	2	4	7
139	8	3	2	5	1	6	7	4
140	8	3	5	2	1	6	4	7
141	8	5	1	4	2	3	7	6
142	8	5	1	4	3	2	6	7
143	8	5	2	3	1	4	7	6
144	8	5	3	2	1	4	6	7

Ces séries d'essais peuvent être utilisées dans plusieurs cas :

- Pour s'affranchir d'une dérive.
- Pour déceler une dérive. S'il n'y a pas d'autres sources d'erreur, les interactions sont comme les nombres (en valeurs absolues) : 0 1 2 4.
- Pour construire un plan 2^{4-1} . L'une des interactions n'étant pas faussée par la dérive, c'est elle que l'on choisira pour alier le quatrième facteur.
- Pour réaliser un blocking et une élimination de la dérive sur un plan 2^4 .

On peut s'affranchir à la fois pour les facteurs principaux :

- d'une variation globale de la moyenne entre deux séries de huit essais (blocking) ;
- d'une dérive linéaire portant sur les 16 essais d'un plan 2^4 . Pour cela on utilise 2 séries *miroir* de 8 essais chacune, par exemple : 64713528 et 82531746.

Voici un exemple illustrant cette technique :

Séries miroir	Facteur (1)	Facteur (2)	Facteur (3)	Facteur (4)	Blocking sur 134
6	+	-	+	-	-
4	+	+	-	+	-
7	-	+	+	+	-
1	-	-	-	-	-
3	-	+	-	-	-
5	-	-	+	+	-
2	+	-	-	+	-
8	+	+	+	-	-
8	+	+	+	+	+
2	+	-	-	-	+
5	-	-	+	-	+
3	-	+	-	+	+
1	-	-	-	+	+
7	-	+	+	-	+
4	+	+	-	-	+
6	+	-	+	+	+

D • INTRODUCTION AU CALCUL MATRICIEL

Les matrices sont des tableaux qui permettent de faire très simplement des calculs très complexes. Les calculs étant réalisés par une machine, les utilisateurs doivent simplement être capables de comprendre la signification des opérations et d'interpréter les résultats.

■ Définitions

□ Matrice

Une *matrice* est un tableau entre deux crochets constitué d'éléments disposés en lignes et en colonnes :

$$\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}$$

Les éléments sont : a, b, c, d, e et f .

La première ligne est : a, b, c .

La deuxième ligne est : d, e, f .

La première colonne est : a, d .

La deuxième colonne est : b, e .

La troisième colonne est : c, f .

□ Ordre d'une matrice

Une matrice d'ordre m par n est un tableau de m lignes et de n colonnes. La matrice précédente est une matrice $(2, 3)$. On indique d'abord le nombre de lignes, puis, séparé par une virgule, le nombre de colonnes. Les deux chiffres sont placés entre parenthèses.

□ Matrice carrée

C'est un tableau dans lequel le nombre de lignes est égal au nombre de colonnes.

□ Matrice ligne

C'est une matrice n'ayant qu'une ligne : $(1, n)$.

□ Matrice colonne

C'est une matrice n'ayant qu'une colonne : $(m, 1)$.

Notation des éléments

Les numéros de ligne et de colonne seront indiqués en indice. Par exemple, la matrice précédente s'écrira :

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

Éléments homologues

Ce sont les éléments qui occupent la même place dans deux matrices de même ordre.

Nature des éléments

Les éléments d'une matrice peuvent être de nature très différente : nombres réels, nombres complexes, polynômes, fonctions, opérateurs et même matrices.

■ Définitions concernant les matrices carrées

Diagonale principale

La diagonale principale d'une matrice carrée est formée de l'ensemble des éléments a_{ii} dont le numéro de ligne est égal au numéro de colonne. Cette diagonale principale est signalée en gras dans la matrice suivante :

$$\begin{bmatrix} \mathbf{a_{11}} & a_{12} & a_{13} \\ a_{21} & \mathbf{a_{22}} & a_{23} \\ a_{31} & a_{32} & \mathbf{a_{33}} \end{bmatrix}$$

Matrice symétrique

C'est une symétrie par rapport à la diagonale principale. Les éléments a_{ij} sont égaux aux éléments a_{ji} :

$$a_{ij} = a_{ji}$$

Exemple :

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 4 & 1 \end{bmatrix}$$

Matrice antisymétrique

Les éléments a_{ij} sont égaux aux éléments a_{ji} changés de signe :

$$a_{ij} = -a_{ji}$$

Exemple :

$$\begin{bmatrix} 0 & 2 & 3 \\ -2 & 0 & 4 \\ -3 & -4 & 0 \end{bmatrix}$$

Matrice unité

C'est une matrice carrée dont tous les éléments sont nuls sauf ceux de la diagonale principale qui sont égaux à 1 :

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Déterminant d'une matrice

Le déterminant d'une matrice est un nombre. Il ne faut pas confondre matrice et déterminant même s'ils se ressemblent beaucoup. Une matrice est un tableau constitué d'éléments. Un déterminant se présente comme un tableau mais que l'on transforme en un nombre à l'aide de règles de calcul bien définies. Dans le domaine des plans d'expériences, les déterminants interviennent dans le calcul de la matrice inverse et dans la construction des plans D-optimaux.

Trace d'une matrice

La trace d'une matrice est la somme de ses éléments diagonaux.

Matrice diagonale

Une matrice diagonale a tous ses éléments égaux à 0 sauf ceux de la diagonale principale.

■ **Opérations sur les matrices**

Égalité de deux matrices

Deux matrices de même ordre sont égales si les éléments homologues sont égaux.

Addition de deux matrices

On ne peut additionner que des matrices de même ordre. On obtient une matrice de même ordre. Les éléments de la matrice somme sont les sommes des éléments homologues des matrices que l'on additionne.

Par exemple, soit deux matrices

$$\mathbf{M}_1 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \quad \text{et} \quad \mathbf{M}_2 = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}$$

Leur somme est :

$$\mathbf{M}_1 + \mathbf{M}_2 = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & a_{13} + b_{13} \\ a_{21} + b_{21} & a_{22} + b_{22} & a_{23} + b_{23} \end{bmatrix} = \mathbf{M}_3$$

La matrice \mathbf{M}_3 est la somme des matrices \mathbf{M}_1 et \mathbf{M}_2 :

$$\mathbf{M}_1 + \mathbf{M}_2 = \mathbf{M}_3$$

Soustraction

On ne peut soustraire que des matrices de même ordre. On obtient une matrice de même ordre. Les éléments de la matrice différence sont les différences des éléments homologues des matrices que l'on soustrait :

$$\mathbf{M}_1 - \mathbf{M}_2 = \begin{bmatrix} a_{11} - b_{11} & a_{12} - b_{12} & a_{13} - b_{13} \\ a_{21} - b_{21} & a_{22} - b_{22} & a_{23} - b_{23} \end{bmatrix}$$

Matrice zéro

La matrice zéro a tous ses éléments nuls.

Multiplication d'une matrice par un scalaire

Tous les éléments de la matrice sont multipliés par ce scalaire :

$$\lambda \times \mathbf{M}_1 = \begin{bmatrix} \lambda a_{11} & \lambda a_{12} & \lambda a_{13} \\ \lambda a_{21} & \lambda a_{22} & \lambda a_{23} \end{bmatrix}$$

Multiplication d'une matrice par une matrice

C'est une opération importante. La multiplication d'une matrice **A** par une matrice **B** donne une matrice **C** :

$$\mathbf{A B} = \mathbf{C}$$

La multiplication n'est possible que si le nombre de colonnes de **A** est égal au nombre de lignes de **B**. Si l'ordre de **A** est (m, p) et celui de **B**, (p, n) , l'ordre de **C** sera (m, n) .

La multiplication de deux matrices se fait **ligne par colonne**, de telle manière que l'élément c_{ij} de la matrice produit soit égal à :

$$c_{ij} = a_{i1} b_{1j} + a_{i2} b_{2j} + \dots + a_{ip} b_{pj}$$

La multiplication matricielle n'est pas commutative :

$$\mathbf{A B} \neq \mathbf{B A}$$

Matrice inverse

Seules les matrices carrées ont des matrices inverses.

La matrice **B** est inverse de la matrice **A** si leur produit est égal à la matrice unité :

$$\mathbf{A B} = \mathbf{I}$$

La matrice **B** inverse de **A**, se note \mathbf{A}^{-1} . La matrice \mathbf{A}^{-1} existe si le déterminant de **A** n'est pas nul. Si le déterminant de **A** est nul, **A** est dite *singulière* ou *dégénérée*.

Transposition d'une matrice

La matrice transposée **A'** de **A** s'obtient en écrivant en colonnes les lignes de **A**. La première ligne de **A** devient la première colonne de **A'**. La deuxième ligne de **A** devient la deuxième colonne de **A'**, etc.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

$$A' = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{bmatrix}$$

■ Algèbre sur les matrices

À partir des définitions et des opérations qui viennent d'être indiquées, il est possible de construire une algèbre des matrices.

□ Transposée d'un produit

$$\begin{aligned} (X_1 X_2)' &= X_2' X_1' \\ (X_1 X_2 X_3)' &= X_3' X_2' X_1' \end{aligned}$$

□ Inverse d'un produit

$$\begin{aligned} (X_1 X_2)^{-1} &= X_2^{-1} X_1^{-1} \\ (X_1 X_2 X_3)^{-1} &= X_3^{-1} X_2^{-1} X_1^{-1} \end{aligned}$$

□ Transposée d'une transposée

$$(X')' = X$$

□ Inverse d'une inverse

$$(X^{-1})^{-1} = X$$

■ Matrices particulières

□ Matrices orthogonales

Une matrice est orthogonale si les produits scalaires de ses colonnes sont tous nuls. La transposée d'une matrice orthogonale est égale à son inverse :

$$X' = X^{-1}$$

□ Matrices d'Hadamard

Ce sont des matrices carrées dont les éléments sont +1 ou -1 et telles que :

$$X' X = n I$$

Elles satisfont aux relations :

$$X' X = X X'$$

$$X' = n X^{-1}$$

$$X^{-1} = \frac{1}{n} X'$$

Exemples :

$$H_2 = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

$$\mathbf{H}_4 = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

Ces matrices \mathbf{H}_n existent pour toute valeur de n multiple de 4, soit 8, 12, 16, 20... Les matrices d'Hadamard sont les matrices utilisées pour construire les plans d'expériences factoriels complets, les plans factoriels fractionnaires, les plans de Plackett et Burman, les tables de Taguchi et la partie factorielle des plans composites.

BIBLIOGRAPHIE

■ Lecture conseillée

BOX G.E.P., HUNTER W.G., HUNTER J.S., *Statistics for Experimenters*, John Wiley, New York, 2^e édition, 2005.

BOX G.E.P., BEHNKEN D.W., Some new three level designs for the study of quantitative variables, *Technometrics*, vol. 2, 1960, p. 455-475.

CORNELL J.A., *Experiment with Mixtures*, John Wiley, New York, 1981.

DANIEL C., Use of half-normal plots in interpreting factorial two-level experiments, *Technometrics*, vol. 1, 1959, p. 311-341.

DERRINGER D., SUICH R., Simultaneous optimisation of several response variables, *J. of Quality Technology*, 12, 1980, p. 214-219.

DOEHLERT D.H., Uniform Shell Design, *Appl. Stat.*, n° 19, 1970, p. 231.

DRAPER N.R., JOHN J.A., Responses Surfaces Designs for Quantitative and Qualitative Variables, *Technometrics*, vol. 30, n° 4, novembre 1988.

DRAPER N.R., SCHMITT H., *Applied Regression Analysis*, John Wiley, New York, 1981.

FISHER R.A., *Statistical Methods for Research Workers*, Oliver and Boyd, Édimbourg.

FISHER R.A., *The Design of Experiments*, Oliver and Boyd, Édimbourg, 1935.

FREUND R., LITTLE R., CREIGHTON L., *Regression using JMP*, SAS Publishing, 2003.

GOUPY J., *Methods for Experimental Design – Principles and Applications for Physicists and Chemists*, Elsevier, Amsterdam, 1993.

GOUPY J., *Plans d'expériences pour surfaces de réponse*, Dunod, Paris, 1999.

GOUPY J., *Plans d'expériences : les mélanges*, Dunod, Paris, 2000.

HOTELLING H., Some problems in weighing and other experimental techniques, *Am. Math. Stat.*, vol. 15, n° 4, 1944, p. 297-306.

MITCHELL T.J., An algorithm for Construction of D-Optimal Experimental Designs, *Technometrics*, vol. 16, n° 2, 1974.

PLACKETT R.L., BURMAN J.P., The design of optimum multifactorial experiments, *Biometrika*, n° 33, 1946.

SALL J., CREIGHTON L., LEHMAN A., *JMP Start Statistics*, Thomson Brook/Cole, 2005.

YATES F., *Experimental Design*, Selected papers, Darien, Connecticut, Hafner, 1970.

YOU DEN W.J., *Statistical Methods for Chemists*, John Wiley, New York, 1951.

■ Ouvrages et articles généraux

AFNOR ISO 9001, *Systèmes qualité – Modèle pour l'assurance de la qualité en conception/développement, production, installation et soutien après vente*, 1987.

AFNOR ISO 9002, *Systèmes qualité – Modèle pour l'assurance de la qualité en production et installation*, 1987.

AFNOR ISO 9003, *Systèmes qualité – Modèle pour l'assurance de la qualité en contrôle et essais finals*, 1987.

AFNOR X 02-110, *Symboles et vocabulaire du calcul matriciel*, juin 1982.

AFNOR X 06-080, *Application de la statistique – Plan d'expériences – Vocabulaire et indications générales*, novembre 1989.

ALEXIS J., *Pratique industrielle de la méthode Taguchi : les plans d'expériences*, AFNOR, 1995.

ATKINSON A.C., The Usefulness of Optimum Experimental Designs, *J. of the Royal Statistical Society B*, vol. 58, n° 1, 1996, p. 59-76.

BALEO J.-N., BOURGES B., COURCOUX P., FAUR-BRASQUET C., LE CLOIREC P., *Méthodologie expérimentale*, Tec & Doc, Paris, 2003.

BENOIST D., TOURBIER Y., GERMAIN-TOURBIER S., *Plans d'expériences : construction et analyse*, Lavoisier – Tec & Doc, Paris, 1994.

BLAQUIÈRE A., *Calcul matriciel* (2 tomes), Hachette, Paris, 1960.

BOX G.E.P., DRAPER N.R., *Empirical Model-Building and Response Surface*, John Wiley, New York, 1987.

BOX G.E.P., DRAPER N.R., *Evolutionary Operation : A Statistical Method for Process Improvement*, John Wiley, New York, 1969.

BOX G.E.P., JONES S., Split-plot designs for robust product experimentation, *J. of Applied Statistics*, 1992, p. 3-26.

BRERETON R.G., Multilevel multifactor designs for multivariate calibration, *Analyst*, vol. 122, n° 12, 1997, p. 1521-1529.

CARLSON R., Preludes to a screening experiment : A tutorial, *Chemometrics and Intelligent Laboratory Systems*, 14, 1992, p. 103-114.

CARLSON R., *Design and Optimisation in organic synthesis*, Elsevier, Amsterdam, 1992.

CORNELL J.A., DENG J.C., Combining Process Variables and Ingredient Components in Mixing Experiments, *J. of Food Science*, vol. 47, 1982, p. 836-843 et 848.

- CORNELL J.A., Embedding Mixture Experiments Inside Factorial Experiments, *J. of Quality Technology*, vol. 23, n° 4, 1990, p. 265-276.
- CORNELL J.A., Analysing Data from Mixture Experiments containing Process Variables : A split-Plot Approach, *J. of Quality Technology*, vol. 20, n° 1, 1988, p. 2-23.
- CZITROM V., Mixture Experiments with Process Variables : D-Optimal Orthogonal Experimental Designs, *Communications in Statistics, Theory and Methods*, vol. 17, n° 1, 1988, p. 105-121.
- DAGNELIE P., *Principes d'expérimentation*, Presses agronomiques de Gembloux, 2003.
- DANIEL C., *Applications of Statistics to Industrial Experimentation*, John Wiley, New York, 1976.
- DRAPER N.R., HERZBERG, On lack of fit, *Technometrics*, vol. 13, 1971, p. 231-241.
- DRAPER N.R., Small composite Designs, *Technometrics*, vol. 27, 1985, p. 173-180.
- DROESBEKE J.-J., FINE J., SAPORTA G., *Plans d'expériences : applications à l'entreprise*, Technip, 1997.
- DUINEVELD C.A.A., SMILDE A.K., DOORNBOS D.A., Comparison of Experimental designs combining Process and Mixture variables, Part I : Design construction and theoretical evaluation et Part II : Design and evaluation on measured data, *Chem. Intel. Lab. Syst.*, vol. 19, 1993, p. 295-308 et 309-318.
- FEDOROV V.V., *Theory of Optimal Experiments*, Academic Press, New York, 1972.
- GENETAY M., *La coloration des matériaux*, Hermes, Paris, 2000.
- GOUPY J., Chemometrics : Experimental Designs, *Encyclopedia of Analytical Sciences*, Academic Press, Londres, 1995, p. 659-666.
- GOUPY J., Boolean Experimental Designs, *Analisis*, vol. 28, n° 7, 2000, p. 563-570.
- GOUPY J.L., Unconventional experimental designs : Theory and Application, *Chem. Intel. Lab. Syst.*, vol. 33, 1996, p. 3-16.
- GOUPY J., Outliers and Experimental Designs, *Chem. Intel. Lab. Syst.*, vol. 35, 1996, p. 145-156.
- GOUPY J., Plans d'expériences, *Traité Analyse chimique et caractérisation*, Techniques de l'ingénieur P230, 1992, p. 1-20.
- GOUPY J., Erreur de dérive et choix de l'ordre des essais d'un plan d'expériences factoriel, *Revue de statistique appliquée*, vol. 37, n° 1, 1989, p. 5-22.
- GOUPY J., Plans d'expériences non conventionnels : théorie et applications (ou comment sauver un plan raté), *Analisis*, 23, 1995, p. 152-158.
- GOUPY J., Modélisation par les plans d'expériences, *Traité Mesures et contrôle*, Techniques de l'ingénieur R275, 2000, p. 1-23.
- GORMAN J.W., HINMAN J.E., Simplex-Lattice Designs for multicomponent systems, *Technometrics*, n° 4, 1962, p. 463-488.

- ISO 5725-2, *Accuracy (trueness and precision) of measurement methods and results, partie 2 : Basic method for the determination of repeatability and reproducibility of a standard measurement method*, 1994.
- ISO 5725-3, *Accuracy (trueness and precision) of measurement methods and results, partie 3 : Intermediate measures of the precision of a standard measurement method*, 1994.
- ISO 3534-3, *Statistique : vocabulaire et symboles, partie 3 : Plans d'expériences*, 1998.
- KHURI A.I., CORNELL J.A., *Response surfaces : Design and Analyses*, Marcel Dekker, New York, 1996.
- LENTH R.V., Quick and Easy Analysis of unreplicated Fractional Factorials, *Technometrics*, vol. 31, n° 4, 1989, p. 469-473.
- LIANG Y.Z., FANG K.T, XU Q.S., Uniform design and its applications in chemistry and chemical engineering, *Chemometrics and Intelligent Laboratory Systems*, vol. 58, n° 1, 2001, p. 43-57.
- LUNDSTEDT T., SEIFERT E., ABRAMO L., THELIN B., NYSTRÖM Å., PETTERSEN J., BERGMAN R., Experimental design and optimisation, *Chemometrics and Intelligent Laboratory Systems*, 42, 1998, p. 3-40.
- MCLEAN R.A., ANDERSON V.L., Extreme vertices design of mixture experiments, *Technometrics*, 8, 1966, p. 447-454.
- MEAD R., *The Design of Experiments : Statistical Principles for Practical application*, Cambridge University Press, Cambridge, 1988.
- MEE R.W., BATES R.L., Split-lot designs : Experiments for multistage batch processes, *Technometrics*, vol. 40, n° 2, 1998, p. 127-140.
- MONTGOMERY D.C., *Design and Analysis of Experiments*, John Wiley, New York, 1984.
- MORGAN E.D., *Chemometrics : Experimental Design*, John Wiley, New York, 1995.
- MYERS R.H., KHURI A.I., CARTER W.H., Response Surface Methodology : 1966-1988, *Technometrics*, vol. 31, n° 2, 1989.
- PIEPEL G.F., CORNELL J.A., Mixture Experiments Approach : Examples, Discussion, and Recommendations, *J. of Quality Technology*, vol. 26, n° 3, 1994, p. 177-195.
- PILLET M., *Introduction aux plans d'expériences par la méthode Taguchi*, Les Éditions d'organisation, Paris, 1992.
- RECHTSCHAFFNER R.L., Saturated Fractions of $2n$ and 3 Factorial Designs, *Technometrics*, vol. 9, 1967, p. 569-575.
- ROQUEMORE K.G., Hybrid Designs for Quadratic Response Surfaces, *Technometrics*, vol. 18, n° 4, 1976, p. 419-423.
- SADO G., SADO M.-C., *Les plans d'expériences : de l'expérimentation à l'assurance qualité*, AFNOR, 1991.

- SALL J.P., Leverage Plots for General Linear Hypotheses, *American Statistician*, vol. 44, n° 4, 1990, p. 303-315.
- SCHEFFE H., Experiments with mixtures, *J. of the Royal Statistical Society B*, vol. 20, 1958, p. 344-360.
- SCHEFFE H., The simplex centroid design for Experiments with mixtures, *J. of the Royal Statistical Society B*, vol. 25, 1963, p. 235-263.
- SCIBILIA B., KOBİ A., CHASSAGNON R., BARREAU A., Plans complémentaires et plans imbriqués, *Revue de statistique appliquée*, vol. XLVIII, n° 1, 2000.
- SCHIMMERLING P., SISSON J.-C., ZAIDI A., *Pratique des plans d'expériences*, Lavoisier – Tec & Doc, 1998.
- SHELTON J.T., KHURI A.I., CORNELL J.A., Selecting check points for testing lack of fit in response surface models, *Technometrics*, 25, 1983, p. 357-365.
- SILVEY S.D., *Optimal Design*, Chapman and Hall, Londres, 1980.
- SMITH A.T., Comparison of information-yield from different experimental designs used in algal toxicity testing of chemical mixtures, *Environmental Pollution*, 102, 1998, p. 205-212.
- SNEE R.D., Experimenting with mixtures, *Chemtech*, 9, 1979, p. 702-710.
- SNEE R.D., Developing Blending Models for Gasoline and Other Mixtures, *Technometrics*, 23, n° 2, 1981, p. 119-130.
- STANSBURY W.F., Development of experimental designs for organic synthetic reactions, *Chemometrics and Intelligent Laboratory Systems*, 36, 1997, p. 199-206.
- STUDENT, The probable error of a mean, *Biometrika*, 6, 1908, p. 1-25.
- SUNDBERG R., Interpretation of unreplicated two-level factorial experiments by examples, *Chem. Intel. Lab. Syst.*, vol. 24, 1994, p. 1-17.
- TOMASSONE R., AUDRAIN S., LESQUOY-TURCKHEIM E., MILIER C., *La régression, nouveaux regards sur une ancienne méthode statistique*, Masson, Paris, 1992.
- VANDER HEYDEN Y., QUESTIER F., MASSART D.L., Ruggedness testing of chromatographic methods : selection of factors and levels, *J. of Pharmaceutical and Biomedical Analysis*, vol. 18, 1998, p. 43-56.
- VANDER HEYDEN Y., QUESTIER F., MASSART D.L., A ruggedness test strategy for procedure related factors : experimental set-up and interpretation, *J. of Pharmaceutical and Biomedical Analysis*, vol. 17, 1998, p. 153-168.
- VANDER HEYDEN Y., BOURGEOIS A., MASSART D.L., Influence of the sequence of experiments in a ruggedness test when drift occurs, *Analytica Chimica Acta*, 347, 1997, p. 369-384.
- VANDER HEYDEN Y., LUYPAERT K., HARTMANN C., MASSART D.L., HOOGMARTENS J., DE BEER J., Ruggedness tests on the HPLC assay of the United States Pharmacopeia XXII for tetracycline hydrochloride : A comparison of experimental designs and statistical interpretation, *Analytica Chimica Acta*, 312, 1995, p. 245-262.

WOLTERS R., KATEMAN G., The construction of simultaneous optimal experimental designs for several polynomials in the calibration of analytical methods, *J. of Chemometrics*, 4, 1990, p. 171-185.

YOUNDEN W.J., *Statistical Methods for Chemists*, John Wiley, New York, 1951.

YUZHU H., Program for Doehlert matrix design of Experiment, *Tr. Anal. Chem.*, vol. 8, n° 4, 1989.

■ Application des plans d'expériences

AZUBEL M., FERNANDEZ F.M., TUDINO M.B, TROCCOLI O.E., Novel application and comparison of multivariate calibration for the simultaneous determination of Cu, Zn and Mn at trace levels using flow injection diode array spectrophotometry, *Analytica Chimica Acta*, vol. 398, n° 1, 1999, p. 93-102.

DORTHE A.M., RAMBERTI J.L., THIENPON T., Experimental designs optimization of chromatographic separation for polycyclic aromatic hydrocarbons in vegetable oils, *Analisis*, vol. 28, n° 7, 2000, p. 587591.

FOURNIER J.-L., EL HOURCH M., TAGLIONI J.-P., FOURNIER J., Application des plans d'expériences à l'analyse de pesticides dans l'eau par chromatographie en phase gazeuse couplée à la spectrométrie de masse, *Analisis*, vol. 26, n° 8, 1998, p. M44-M52.

GUERVENOU J., GIAMARCHI P., COULOUARN C., GUERDA M., LE LEZ C., OBOYET T., Experimental design methodology and data analysis technique applied to optimise an organic synthesis, *Chem. Intel. Lab. Syst.*, vol. 63, 2002, p. 81-89.

GUNDUZ U., Partitioning of bovine serum albumin in an aqueous two-phase system : optimisation of partition coefficient, *J. Chromatogr. B : Biomed. Applied*, vol. 743, n° 1-2, 2000, p. 259-262.

HECK G., MILEHAM C., MARTIN G.J., Hydrogen exchange in aromatic compounds : substituent effects studied by experimental designs, *Analisis*, vol. 25, 1997, p. 202-206.

MASSUMI A., NAJAFI N.M., BARZEGARI H., Speciation of Cr(VI)/Cr(III) in environmental waters by fluorimetric method using central composite, full and fractional factorial design, *Microchemical Journal*, 72, 2002, p. 93-101.

MORAND P., VAULLERIN M., ESPAGNACQ A., MORIN-ALLORY L., Explosive formulation by experimental design, *Analisis*, vol. 26, 1998, p. 291-294.

MORITA K., KUBOTA K., AISHIMA T., Sensory characteristics and volatile components in aromas of boiled prawns prepared according to experimental designs, *Food Research International*, vol. 34, n° 6, 2001, p. 473-481.

MATHIA P., OUADOU M., LANTERI P., LONGERAY R., CLAUDAUD B., Optimisation design in abrasive resistance of refractory concretes, *Applied Mechanics and Engineering*, vol. 1, n° 3, 1996, p. 367-380.

MORITA K., KUBOTA K., AISHIMA T., Investigating influence of pH and parts on sensory characteristics and volatile components in boiled squid using experimental designs, *J. of Food Science*, vol. 67, n° 2, 2002, p. 848-854.

MOTAMED B., BOHM J.L., HENNEQUIN D., TEXIER H., MOSRATI R., BARILLIER D., Development of an HPLC method for the determination of phenolic by-products : optimisation of the separation by means of the experimental designs methodology, *Analysis*, vol. 28, n° 7, 2000, p. 592-599.

ROGOWSKI I., TABONI A., LONGERAY R., MIDOL A., LANTERI P., *Les plans d'expériences pour étudier l'usure du matériel sportif : application à l'étude de la perte de tension de cordage de raquettes de tennis*, Chimométrie 2000, Paris, 2000.

RUDAZ S., CHERKAOUI S., GAUVRIT J.-Y., LANTERI P., VEUTHEY J.-L., Experimental designs to investigate capillary electrophoresis-electrospray ionization-mass spectrometry enantioseparation with the partial-filling technique, *Electrophoresis*, vol. 22, n° 15, 2001, p. 3316-3326.

VEUTHEY J.-L., LANTERI P.-E., LONGERAY J.-R., GAUVRIT J.-Y., VAREGIO E., Central composite design in the chiral analysis of amphetamines by capillary electrophoresis, *Electrophoresis*, vol. 18, 1997, p. 931-937.

XU Q.S., LIANG Y.Z., FANG K.T., The effects of different experimental designs on parameter estimation in the kinetics of a reversible chemical reaction, *Chemometrics and Intelligent Laboratory Systems*, vol. 52, n° 2, 2000, p. 155-166.

■ Principaux logiciels pour plans d'expériences

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

A

α 188
alias 97, 104
analyse
 de la variance 82, 256, 263, 290
 des résidus 89
Augmented Simplex-centroid designs 214

B

blocking 149, 151, 161, 305
Box (calcul de) 99, 102
Behnken (plan de) 180, 194

C

calcul de Box 99, 102
carrés
 de Youden 255, 266
 gréco-latins 265
 latins 255, 264
 moyens des résidus 84
coefficients 13, 81
 de détermination R^2 85
concomitance 104
confusion 104
contraste 97, 104
custom designs 221

D

Daniel (diagramme de) 121
degrés de liberté 74, 83
dérive 149
désirabilité 60
diagramme
 de Daniel 121
 de Pareto 121
 des effets 39
 des interactions 39
distribution 67, 74
 de Laplace-Gauss 74
 normale 74

Doehlert (plan de) 181, 199
domaine 273
 d'étude 6
 d'un facteur 6
D-optimalité 225
droite de Henry 121

E

écarts 83
 à la moyenne 82, 83
écart-type 67, 71, 72
échantillon 67, 72
effet du facteur 22
erreurs expérimentales 13, 71, 148, 275
espace expérimental 5
expériences de contrôle 46
extrapolation maîtrisée 66

F

F de Fisher 84
facteurs 4, 81, 273
 booléens 4
 continus 4
 contrôlés 34, 148
 de blocking 152
 de procédé 241
 discrets 4, 255
 indépendants 207
 influent 121
 ordonnables 4
formule de passage 8

G

générateur d'alias 104
GGA 110

H

Hadamard (matrice de) 142, 188, 311
Henry (droite de) 121
histogramme 73
hypothèses d'interprétation 98

I

interaction 27
 intervalle de confiance des réponses 79
 isovariance par rotation 188

L-M

lack of fit 13, 81
 Laplace-Gauss (distribution de) 74
 manque d'ajustement 13, 81
 matrice
 d'expériences 17
 d'expérimentation 17
 d'Hadamard 142, 188, 311
 d'information 14
 de calcul des coefficients 13
 de dispersion 14
 du modèle 13
 orthogonale 311
 modèle
 des plans de mélanges 215
 du second degré 179
 PDAI 21, 35, 54
 postulé 20
 modélisation mathématique 12, 215, 255
 moyenne 67, 71

N

niveau 6

O

optimalité 225
 ordre de Yates 53, 283
 orthogonalité 142, 188, 311

P

Pareto (diagramme de) 121
 Pas 8
 Plackett et Burman (plan de) 142
 plan
 complémentaire 133, 134
 composite 179, 184
 d'expériences 17
 de base 102
 de Box-Behnken 180, 194
 de Doehlert 181, 199
 de mélange 207
 centré 214
 augmenté 214
 de Plackett et Burman 142
 de Scheffé 212
 en étoile 188

en réseaux 212
 factoriel
 complet 49, 142
 fractionnaire 93
 gréco-latin 255
 saturé 115
 sur mesure 221

points

au centre 70, 188, 214, 276
 de contrôle 70, 147
 de mélange 208
 en étoile 179, 188
 expérimentaux 9

population 67, 72

pure error 13

p-value 88

R

R carré 85
 randomisation 150, 175, 274
 relation
 d'équivalence 102
 de définition 111
 réponse 4, 273
 corrigée de la moyenne 82
 résidus 82, 83
 RMSE 84

S

Scheffé (plan de) 212
Simplex lattice designs 212
Simplex-centroid designs 214
 surface de réponse 12, 179
 système d'équations 13

T

tableau d'expérimentation 17
 transmission des erreurs 78

V

variables centrées réduites 8
 variance 72, 78
 vecteur
 des coefficients 13
 des écarts 13
 des réponses 13

Y

Yates 53, 161, 283
 Youden (carrés de) 255, 266

049744 - (I) - (0,8) - CSBM90° - TYP - CDD
Imprimerie CHIRAT - 42540 Saint-Just-la-Pendue
Dépôt légal : Novembre 2006 - N° 2418
Dépôt légal de la 2^e édition : Juin 2001

Imprimé en France

INSTRUCTIONS POUR UTILISER LE LOGICIEL JMP®

Avant l'installation du logiciel JMP®, veuillez lire soigneusement ces premières instructions, puis consulter le blog du livre (<http://www.plansdexperiences.com>) dans lequel elles sont largement détaillées.

Veuillez également à bien conserver le numéro qui figure sur le sticker qui accompagne le CD-Rom. Il vous servira de numéro de série (*serial number*) et de numéro d'activation (*activation number*).

Pour utiliser le logiciel, vous devez effectuer les opérations suivantes :

1. Veuillez introduire le CD-Rom dans le lecteur de votre ordinateur. Cliquez sur **Install JMP 6** dans la fenêtre d'installation. La fenêtre d'installation apparaît. Cliquez sur **Install JMP 6** et suivez les instructions. Cette première opération installe le logiciel sur le disque dur de votre ordinateur.

Nota bene

Si une version antérieure de JMP® est déjà installée sur votre ordinateur, l'installateur procède d'abord à une opération de mise à jour. Dans la fenêtre InstallShield Wizard qui apparaît alors, choisissez le bouton radio Remove. Cliquez sur OK dans la fenêtre Confirm Uninstall, puis sur No dans la fenêtre Caution Alert. Pour terminer la maintenance, cliquez sur Finish dans la fenêtre InstallShield Wizard. La fenêtre d'installation réapparaît. Cliquez de nouveau sur Install JMP 6 et suivez les instructions.

2. Vous pouvez alors utiliser immédiatement le logiciel. Ce logiciel fonctionnera pendant un mois.
3. Pour étendre la durée d'utilisation, vous devez activer le logiciel soit immédiatement soit plus tard. Conservez de toute façon le numéro du sticker à portée de main.
4. Une fenêtre vous invite à activer le logiciel. Vous êtes mis en relation avec le site de la société **SAS Institute Inc.** après avoir saisi en face de **Serial Number** le numéro d'activation qui figure sur le sticker accompagnant le CD-Rom. Le logiciel sera alors disponible pendant six mois.
5. Vous trouverez les principales indications pour bien démarrer et utiliser JMP® dans le chapitre 14 de cet ouvrage. Pour compléter vos connaissances, vous pouvez aller sur le blog du livre (<http://www.plansdexperiences.com>). Vous y trouverez un manuel téléchargeable (en français) d'introduction au logiciel.

Afin de réussir l'installation et l'activation, nous vous conseillons de consulter notre blog où nous vous indiquons tous les détails des procédures à suivre.

Vous y trouverez également des explications supplémentaires sur les exemples, de nouveaux exemples et des réponses aux questions posées par les lecteurs du livre, les utilisateurs du logiciel et les expérimentateurs utilisant des plans d'expériences.

Ce CD-Rom d'accompagnement de l'ouvrage de Jacques Goupy et Lee Creighton, *Introduction aux plans d'expériences* (3^e édition, 2006), est offert par DUNOD et SAS Institute Inc. Il ne peut être vendu séparément, ni repris ni échangé.

Jacques Goupy • Lee Creighton

INTRODUCTION AUX PLANS D'EXPÉRIENCES

À travers de nombreux exemples réels empruntés à des expérimentateurs issus de secteurs industriels variés, cet ouvrage expose de façon progressive et appliquée les techniques nécessaires à la conduite d'une étude par plans d'expériences. Cette **3^e édition** a été complétée pour aborder l'ensemble des types de plans :

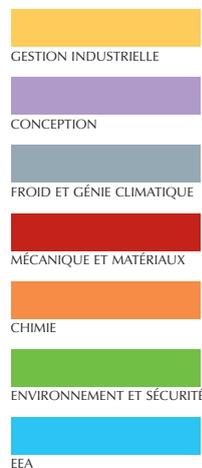
- plan factoriel complet ;
- plan fractionnaire ;
- plan pour surfaces de réponse (composites, de Box-Behnken, de Doehlert) ;
- plan de mélanges ;
- plan optimal ;
- plan pour variables discrètes.

Cet ouvrage constitue un outil précieux pour tous les expérimentateurs en recherche, en développement ou en industrialisation dans de très nombreux domaines (agroalimentaire, chimie, métallurgie, mécanique, électronique, pharmacie, médecine, informatique, etc.).

Le **CD-Rom** gratuit accompagnant l'ouvrage contient une version récente et complète du logiciel de plans d'expériences JMP® (SAS Institute Inc.). Ce logiciel permet de réaliser rapidement les calculs les plus complexes et de tracer les diagrammes des exemples d'applications proposés.



<http://www.plansdexpériences.com>



3^e édition

JACQUES GOUPY

est ingénieur-conseil auprès d'importantes sociétés et enseigne la méthodologie des plans d'expériences en écoles d'ingénieurs et dans l'industrie.

Il est l'auteur de plusieurs livres et articles de référence sur le sujet, dont *Pratiquer les plans d'expériences*, *Plans d'expériences pour surfaces de réponse* et *Plans d'expériences : les mélanges* parus aux éditions Dunod.

LEE CREIGHTON

est professeur de mathématiques et responsable de la documentation JMP® chez SAS Institute Inc. Il est l'auteur de plusieurs ouvrages de statistiques dont *JMP Start Statistics : a Guide to Statistics and Data Analysis* et *Regression with JMP*.

Pour travailler plus efficacement, téléchargez les fichiers JMP® des exemples développés dans l'ouvrage.



ISBN 2 10 049744 8

L'USINE NOUVELLE

www.dunod.com

